

# Comparison of Three Full-Reference Color Image Quality Measures

Eugene GirshTel, Vitaliy Slobodyan, Jonathan S. Weissman, Ahmet M. Eskicioglu<sup>1</sup>  
Department of Computer and Information Science, CUNY Brooklyn College  
2900 Bedford Avenue, Brooklyn, NY 11210, USA

## ABSTRACT

Image quality assessment plays a major role in many image processing applications. Although much effort has been made in recent years towards the development of quantitative measures, the relevant literature does not include many papers that have produced accomplished results. Ideally, a useful measure should be easy to compute, independent of viewing distance, and able to quantify all types of image distortions. In this paper, we will compare three full-reference full-color image quality measures (M-DFT, M-DWT, and M-DCT). Assume the size of a given image is  $n \times n$ . The transform (DFT, DWT, or DCT) is applied to the luminance layer of the original and degraded images. The transform coefficients are then divided into four bands, and the following operations are performed for each band: (a) obtain the magnitudes  $M_{oi}$ ,  $i=1, \dots, (n \times n/4)$  of original transform coefficients, (b) obtain the magnitudes  $M_{di}$ ,  $i=1, \dots, (n \times n/4)$  of degraded transform coefficients, (c) compute the absolute value of the differences:  $|M_{oi} - M_{di}|$ ,  $i=1, \dots, (n \times n/4)$ , and (d) compute the standard deviation of the differences. Finally, the mean of the four standard deviations is obtained to produce a single value representing the overall quality of the degraded image. In our experiments, we have used five degradation types, and five degradation levels. The three proposed full-reference measures outperform the Peak-Signal-to-Noise Ratio (PSNR), and two state-of-the-art metrics Q and MSSIM.

**Keywords:** image quality, quantitative measures, subjective evaluation, PSNR, Q, MSSIM, DFT, DWT, and DCT.

## 1. INTRODUCTION

An important criterion used in the classification of image quality measures is the type of information needed to evaluate the distortion in degraded images. Measures that require both the original image and the distorted image are called "full-reference" or "non-blind" methods, measures that do not require the original image are called "no-reference" or "blind" methods, and measures that require both the distorted image and partial information about the original image are called "reduced-reference" methods.

Although no-reference measures are needed in some applications in which the original image is not available, they can be used to predict only a small number of distortion types. In the current literature, a few papers attempt to predict JPEG compression artifacts [1,2,3,4], and others blurring and JPEG 2000 artifacts [5,6]. Reduced-reference measures are between full-reference and no-reference measures; [7] evaluates the quality of JPEG and JPEG2000 coded images whereas [8] provides assessment for JPEG and JPEG2000 compressed images, images distorted by white Gaussian noise, Gaussian blur, and the transmission errors in JPEG2000 bit streams.

The applicability of full-reference measures is much wider. They can be used to estimate a spectrum of distortions that range from blurriness and blockiness to several types of noise. Recent examples of such measures are given in Table 1.

---

<sup>1</sup> Corresponding author's email address: eskicioglu@sci.brooklyn.cuny.edu

# REPORT DOCUMENTATION PAGE

AFRL-SR-AR-TR-07-0146

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, gathering existing data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE		3. REPORT TYPE AND DATES COVERED Final / 12-1-2005 -11/30/2006	
4. TITLE AND SUBTITLE Quality Measures Using Singular Value Decomposition				5. FUNDING NUMBERS FA9550-05-1-0400	
6. AUTHOR(S) Ahmet Eskicioglu					
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Research Foundation of CUNY 230 West 41 <sup>st</sup> Street, 7 <sup>th</sup> Floor New York, New York, 10036-7296				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) USAF, AFRL AF Office of Scientific Research 875 North Randolph Street, Rm. 3112 Arlington, VA 22203 <i>Dr Robert Herklotz/NM</i>				10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES					
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approve for public release, distribution unlimited.					12b. DISTRIBUTION CODE
13. ABSTRACT (Maximum 200 Words) Image quality assessment plays a major role in many image processing applications. Although much effort has been made in recent years towards the development of quantitative measures, the relevant literature does not include many papers that have produced accomplished results. Ideally, a useful measure should be easy to compute, independent of viewing distance, and able to quantify all types of image distortions. In this paper, we will compare three full-reference full-color image quality measures (M-DFT, M-DWT, and M-DDCT). Assume the size of a given image is nxn. The transform (DFT, DWT, or DCT) is applied to the luminance layer of the original and degraded images. The transform coefficients are then divided into four bands, and the following operations are performed for each band. Finally, the mean of the four standard deviations is obtained to produce a single value representing the overall quality of the degraded image.					
14. SUBJECT TERMS					15. NUMBER OF PAGES
					16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT		18. SECURITY CLASSIFICATION OF THIS PAGE		19. SECURITY CLASSIFICATION OF ABSTRACT	
				20. LIMITATION OF ABSTRACT	

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)  
Prescribed by ANSI Std. Z39-18  
298-102

5/11/07

Table 1. Full-reference image quality measures

Publication	Domain type	Type of distortion predicted
Wang and Bovik [11]	Pixel	Impulsive salt-pepper noise, additive Gaussian noise, multiplicative speckle noise, mean shift, contrast stretching, blurring, and JPEG compression
Wang, Bovik, Sheikh and Simoncelli [12]	Pixel	JPEG compression, JPEG 2000 compression
Van der Weken, Nachtgael and Kerre [9]	Pixel	Salt and pepper noise, enlightening, and darkening
Beghdadi and Pesquet-Popescu [10]	Discrete Wavelet Transform (DWT)	Gaussian noise, grid pattern, JPEG compression

A full-reference paper [11] presents a new numerical measure for gray scale images, called the Universal Image Quality Index, Q, which is defined as

$$Q = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2 + \sigma_y^2)(\mu_x^2 + \mu_y^2)},$$

where  $x_i, y_i, i = 1, \dots, n$ , represent the original and distorted signals, respectively,  $\mu_x = \frac{1}{n} \sum_{i=1}^n x_i$ ,  $\mu_y = \frac{1}{n} \sum_{i=1}^n y_i$ ,

$$\sigma_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)^2, \quad \sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \mu_y)^2, \quad \text{and} \quad \sigma_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y).$$

The dynamic range of Q is [-1,1], with the best value achieved when  $y_i = x_i, i = 1, 2, \dots, n$ . This index models any distortion as a combination of three different factors - loss of correlation, mean distortion, and variance distortion:

$$\frac{\sigma_{xy}}{\sigma_x \sigma_y}, \quad \frac{2\mu_x \mu_y}{(\mu_x)^2 + (\mu_y)^2}, \quad \text{and} \quad \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}.$$

It is applied to 512x512, 8 bits/pixel Lena using a sliding window approach with a window size of 8x8. The index is computed for each window, leading to a quality map of the image. The overall quality index is the average of all the Q values in the quality map:

$$Q = \frac{1}{M} \sum_{j=1}^M Q_j, \quad M = \text{total number of windows.}$$

Q produces unstable results when either  $(\mu_x^2 + \mu_y^2)$  or  $(\sigma_x^2 + \sigma_y^2)$  is very close to zero. To avoid this problem, the measure has been generalized to the Structural Similarity Index (SSIM) [12]:

$$\text{SSIM} = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Q is a special case of SSIM that can be derived by setting  $C_1$  and  $C_2$  to 0. The performance of SSIM has been tested using a database of JPEG and JPEG 2000 compressed color images. In the experiments, only the luminance component of each compressed image is used. The authors argue that the use of color components does not significantly change the performance of the model, even though they acknowledge the fact that this may not be generally true for color image

quality assessment. As in the case of Q, the overall image quality MSSIM is obtained by computing the average of SSIM values over all windows:

$$\text{MSSIM} = \frac{1}{M} \sum_{j=1}^M \text{SSIM}_j$$

In this paper, we will compare three full-reference full color image quality measures:

- M-DFT: A Full-Reference Color Image Quality Measure in the DFT Domain [13]
- M-DWT: A Full-Reference Color Image Quality Measure in the DWT Domain [14]
- M-DCT: A Full-Reference Color Image Quality Measure in the DCT Domain [15]

YUV and RGB are two of the commonly used color models for images and video. The model YUV is a linear transformation between the gamma-corrected RGB components to produce a luminance signal and a pair of chrominance signals. The luminance signal conveys color brightness levels, and each chrominance signal gives the difference between a color and a reference white at the same luminance. A common approach employed in developing a quality measure for color images is to use only the luminance signal.

Assume the size of a given image is  $n \times n$ . Each proposed algorithm is as follows:

2. Apply the transform (DFT, DWT, or DCT) to the luminance layer of the original image.
3. Apply the transform to the luminance layer of the degraded image.
4. Divide the transform coefficients into four bands.
5. For each band, perform the following operations:
  - a. Obtain the magnitudes  $M_{or}, i=1, \dots, (n \times n/4)$  of original transform coefficients.
  - b. Obtain the magnitudes  $M_{di}, i=1, \dots, (n \times n/4)$  of degraded transform coefficients.
  - c. Compute the absolute value of the differences:  $|M_{or} - M_{di}|, i=1, \dots, (n \times n/4)$ .
  - d. Compute the standard deviation of the differences.
6. Obtain the mean of four standard deviations.

## 2. EXPERIMENTS

The three measures were applied to four 512x512 full color images (Lena, Goldhill, Peppers, and Airplane). The images, shown in Figure 1, have different frequencies, ranging from low frequency content (e.g., clouds in the Airplane image) to high frequency content (e.g., feathers in the Lena image).

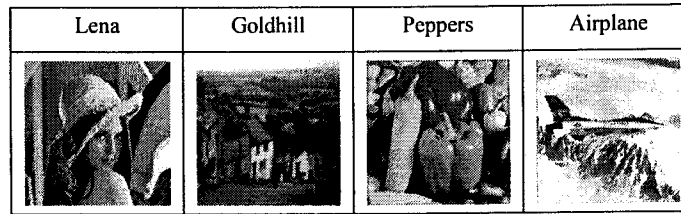


Figure 1. Test images

Table 2 shows the tools and parameters for five degradation types, and five degradation levels. Note that all of these degradations are performed in the pixel domain. The 25 distorted images for each test image are shown in Figures 3-6.

For each test image, high quality printouts of 25 distorted images were subjectively evaluated by approximately 15 observers. The printer was a Hewlett-Packard printer with model number "HP Color Laserjet 4600dn." The 8"x8" images were printed on 8-1/2"x11" white paper with the basis weight 20lb and brightness 84.

Table 2. Distortion types and distortion levels

Type \ Level	Level 1	Level 2	Level 3	Level 4	Level 5
JPEG (XnView)	20:1	40:1	60:1	80:1	100:1
JPEG2000 (XnView)	20:1	40:1	60:1	80:1	100:1
Gaussian blur (Photoshop)	1	2	3	4	5
Gaussian noise (Photoshop)	3	6	9	12	15
Sharpening (XnView)	10	20	30	40	50

The observers were chosen among the graduate students and instructors from the Department of Computer and Information Science at Brooklyn College. About half of the observers were familiar with image processing, and the others had a general computer science background. They were asked to rank the images using a 50-point scale in two ways: within a given distortion type (i.e., rating of the 5 distorted images), and across five distortion types (i.e., rating of the 5 distorted images for each distortion level).

As the proposed measure is not HVS-based, no viewing distance was imposed on the observers in the experiment. In subjective evaluation [16], the widest scale is [0,10]. In order to give the observers a wider scale, grade 1 was assigned to the best image, and grade 50 was assigned to the worst image.

We will show the overall performance of the measures using two criteria: Correlation Coefficient (CC) and Root Mean Squared Error (RMSE) between Mean Opinion Score (MOS) and objective prediction. The real success of objective quality assessment can be determined by predicting the quality not only within a given distortion type but also across different distortion types.

We will also compute two additional sets of data in comparing the performance of the measures:

- CC and RMSE within each of the 5 distortion types, and
- CC and RMSE across each of the 5 distortion levels.

Finally, we will compare the performance of the three measures with PSNR, and two state-of-the-art metrics, Q and MSSIM.

The main purpose of the Video Quality Experts Group (VQEG) is to provide input to the relevant standardization bodies responsible for producing international Recommendations regarding the definition of an objective Video Quality Metric (VQM) in the digital domain.

In the FR-TV Phase I testing and validation, a nonlinear mapping between the objective model outputs and subjective quality ratings was used [17]. The performance of the 9 proponent models was evaluated after compensating for the nonlinearity. In our experiments, we followed the same procedure by fitting a logistic curve to establish a nonlinear mapping. The logistic function has the form

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\alpha x)},$$

where  $\tau$  is a constant parameter.


























Level/Distortion type	Gaussian blur	JPEG	JPEG 2000	Gaussian noise	Sharpening
1					
2					
3					
4					
5					

Figure 3. Lena

























Level/Distortion type	Gaussian blur	JPEG	JPEG 2000	Gaussian noise	Sharpening
1					
2					
3					
4					
5					

Figure 4. Goldhill





















Level/Distortion type	Gaussian blur	JPEG	JPEG 2000	Gaussian noise	Sharpening
1					
2					
3					
4					
5					

Figure 5. Peppers


























Level/Distortion type	Gaussian blur	JPEG	JPEG 2000	Gaussian noise	Sharpening
1					
2					
3					
4					
5					

Figure 6. Airplane

## 2.1 M-DFT: A Full-Reference Color Image Quality Measure in the DFT Domain

Figure 7 shows the curves fitted for all the four measures (PSNR, Q, MSSIM, and M-DFT) compared.

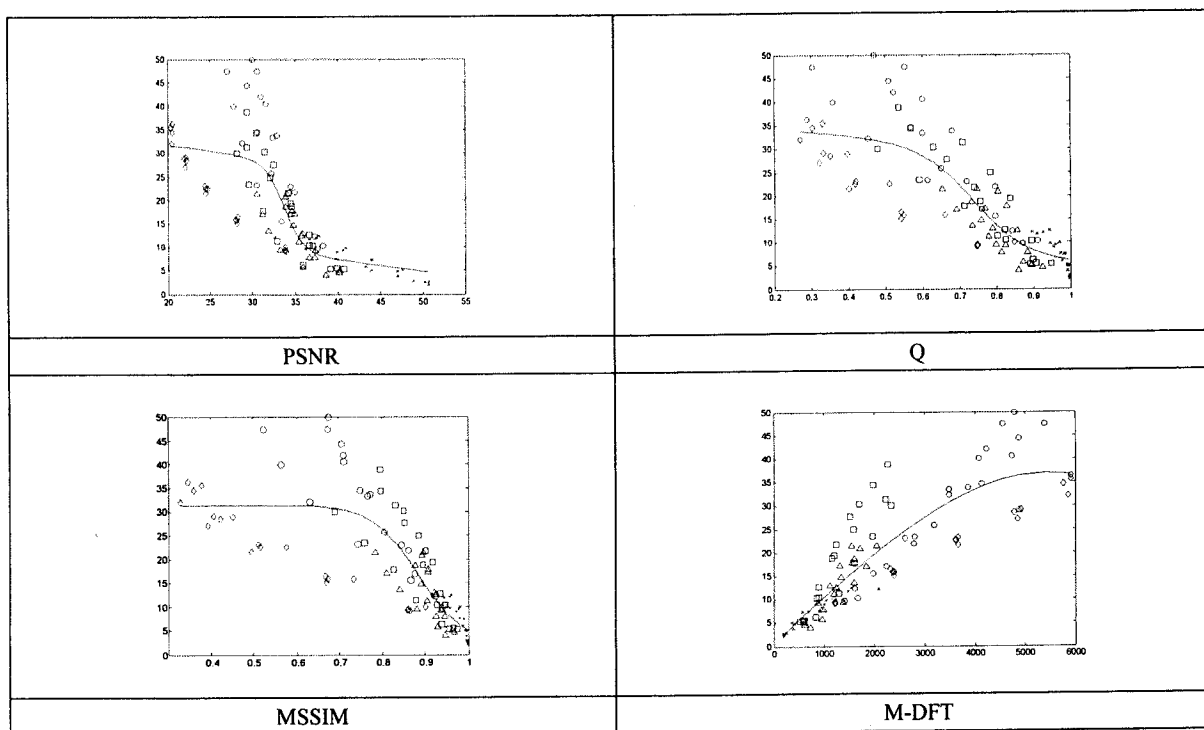


Figure 7. Comparison of the scatter plots for PSNR, Q, MSSIM, and M-DFT. In each plot, the y-axis represents the Mean Opinion Score (MOS), the x-axis represents the quantitative measure, and each mark represents one distorted image. The mapping between the distortion types and the marks is as follows: JPEG ( $\square$ ), JPEG 2000 ( $\Delta$ ), Gaussian blur ( $\circ$ ), Gaussian noise ( $\diamond$ ), Sharpening ( $\times$ ).

Table 3 displays the overall performance of the measures using two criteria: Correlation Coefficient (CC) and Root Mean Squared Error (RMSE) between MOS and objective prediction.

Table 3. Comparison of four measures

Criteria/Measure	PSNR	Q	MSSIM	M-DFT
CC	0.8038	0.8482	0.8282	0.8800
RMSE	7.0857	6.3096	6.6749	5.6579

The performance of the measures within each distortion type and across different distortion types are given in Tables 4 and 5, respectively. We observe that M-DFT outperforms all the other measures not only in overall performance but also within each distortion type and across each distortion level.



Table 4. (a) CC-based performance within each distortion type

Distortion type/Measure	PSNR	Q	MSSIM	M-DFT
JPEG	0.8877	0.9136	0.8768	0.9314
JPEG2000	0.7799	0.7810	0.8354	0.9042
Gaussian blur	0.8773	0.9124	0.9248	0.9804
Gaussian noise	0.9947	0.9585	0.9766	0.9950
Sharpening	0.9513	0.9662	0.9627	0.9739

(b) RMSE-based performance within each distortion type

Distortion type/Measure	PSNR	Q	MSSIM	M-DFT
JPEG	4.7647	4.2082	4.9762	3.7709
JPEG2000	3.5505	3.5387	3.1151	2.4208
Gaussian blur	6.1557	5.2497	4.8780	2.5280
Gaussian noise	0.9181	2.5556	1.9291	0.8941
Sharpening	1.0498	0.8783	0.9215	0.7723

Table 5. (a) CC-based performance across each distortion level

Distortion level/Measure	PSNR	Q	MSSIM	M-DFT
1	0.8024	0.7783	0.8436	0.9726
2	0.8246	0.8405	0.8542	0.9426
3	0.8361	0.8408	0.8402	0.8710
4	0.8422	0.8601	0.8549	0.8969
5	0.8358	0.8818	0.8755	0.9057

(b) RMSE-based performance across each distortion level

Distortion level/Measure	PSNR	Q	MSSIM	M-DFT
1	2.1151	2.2253	1.9028	0.8239
2	3.3381	3.1974	3.0679	1.9707
3	4.6552	4.5944	4.6026	4.1711
4	5.6156	5.3127	5.4026	4.6057
5	6.7775	5.8218	5.9646	5.2368

## 2.2 M-DWT: A Full-Reference Color Image Quality Measure in the DWT Domain

Figure 8 shows the curves fitted for all the four measures (PSNR, Q, MSSIM, and M-DWT) compared.

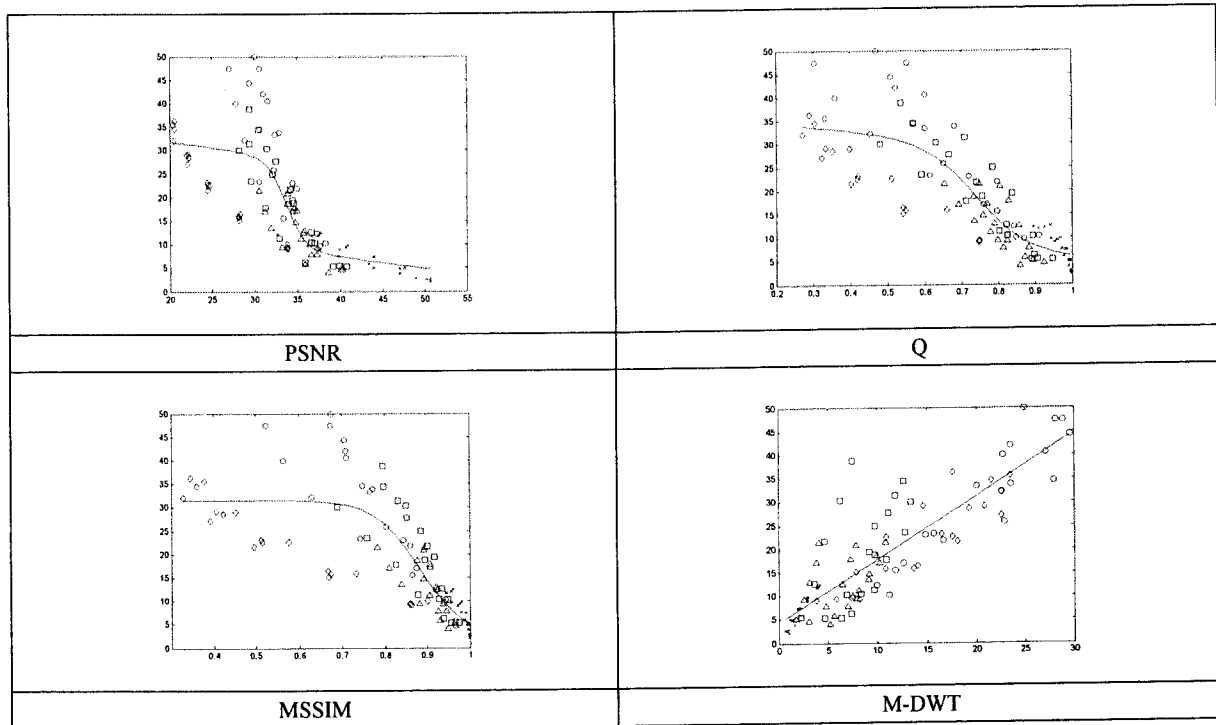


Figure 8. Comparison of the scatter plots for PSNR, Q, MSSIM, and M-DWT. In each plot, the y-axis represents the Mean Opinion Score (MOS), the x-axis represents the quantitative measure, and each mark represents one distorted image. The mapping between the distortion types and the marks is as follows: JPEG ( $\square$ ), JPEG 2000 ( $\Delta$ ), Gaussian blur ( $\circ$ ), Gaussian noise ( $\diamond$ ), Sharpening ( $\times$ ).

Table 6 displays the overall performance of the measures using two criteria: Correlation Coefficient (CC) and Root Mean Squared Error (RMSE) between MOS and objective prediction.

Table 6. Comparison of four measures

Criteria\Measure	PSNR	Q	MSSIM	M-DWT
CC	0.8038	0.8482	0.8282	0.8724
RMSE	7.0857	6.3096	6.6749	5.8214

The performance of the measures within each distortion type and across different distortion types are given in Tables 7 and 8, respectively.

We observe that the overall performance of M-DWT is better than the other three measures. The performance is not as good as the performance of PSNR, Q, and MSSIM for three types of distortion. However, the performance of M-DWT is considerably more consistent across distortion levels, a problem which is very difficult to solve.

Table 7. (a) CC-based performance within each distortion type

Distortion type\Measure	PSNR	Q	MSSIM	M-DWT
JPEG	0.8877	0.9136	0.8768	0.6123
JPEG2000	0.7799	0.7810	0.8354	0.5559
Gaussian blur	0.8773	0.9124	0.9248	0.9336
Gaussian noise	0.9947	0.9585	0.9766	0.8834
Sharpening	0.9513	0.9662	0.9627	0.9777

(b) RMSE-based performance within each distortion type

Distortion type\Measure	PSNR	Q	MSSIM	M-DWT
JPEG	4.7647	4.2082	4.9762	8.1811
JPEG2000	3.5505	3.5387	3.1151	4.7102
Gaussian blur	6.1557	5.2497	4.8780	4.5941
Gaussian noise	0.9181	2.5556	1.9291	4.2007
Sharpening	1.0498	0.8783	0.9215	0.7150

Table 8. (a) CC-based performance across each distortion level

Distortion level\Measure	PSNR	Q	MSSIM	M-DWT
1	0.8024	0.7783	0.8436	0.8788
2	0.8246	0.8405	0.8542	0.9417
3	0.8361	0.8408	0.8402	0.9010
4	0.8422	0.8601	0.8549	0.8979
5	0.8358	0.8818	0.8755	0.8963

(b) RMSE-based performance across each distortion level

Distortion level\Measure	PSNR	Q	MSSIM	M-DWT
1	2.1151	2.2253	1.9028	1.6911
2	3.3381	3.1974	3.0679	1.9851
3	4.6552	4.5944	4.6026	3.6815
4	5.6156	5.3127	5.4026	4.5840
5	6.7775	5.8218	5.9646	5.4741

### 2.3 M-DCT: A Full-Reference Color Image Quality Measure in the DCT Domain

Figure 9 shows the curves fitted for all the four measures (PSNR, Q, MSSIM, and M-DCT) compared.

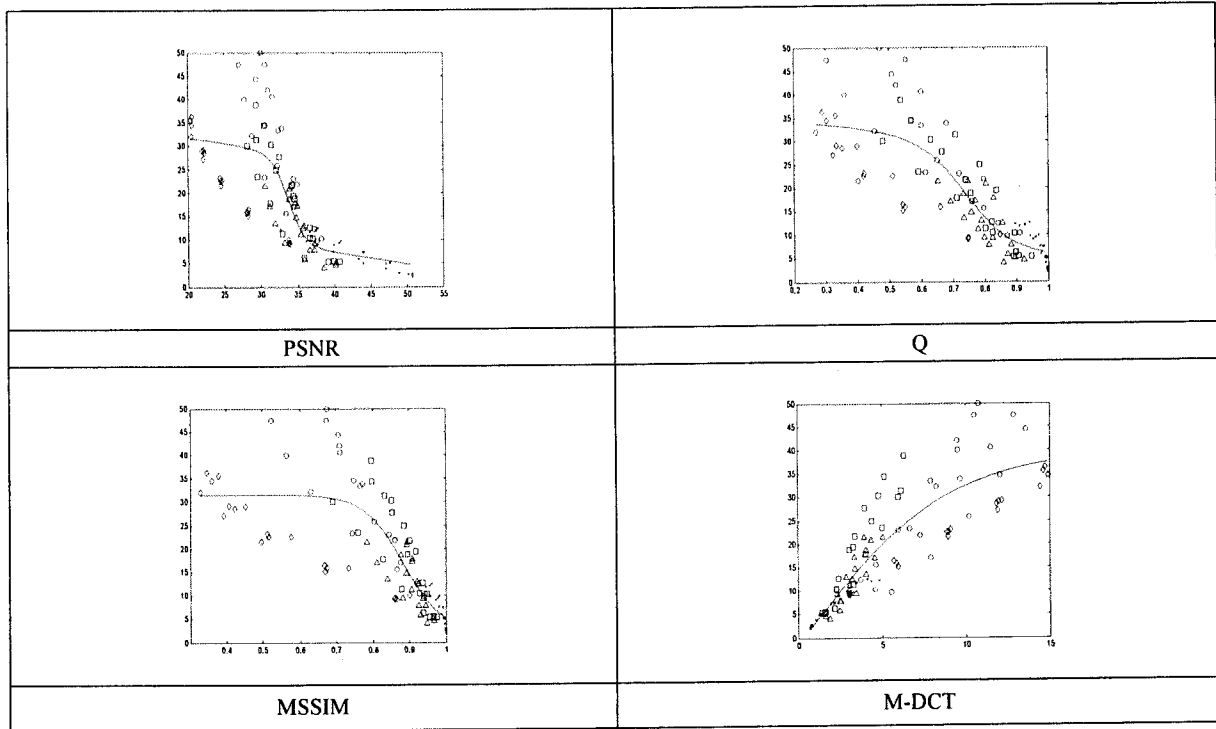


Figure 9. Comparison of the scatter plots for M-DCT, PSNR, Q, and MSSIM. In each plot, the y-axis represents the Mean Opinion Score (MOS), the x-axis represents the quantitative measure, and each mark represents one distorted image. The mapping between the distortion types and the marks is as follows: JPEG ( $\square$ ), JPEG 2000 ( $\Delta$ ), Gaussian blur ( $\circ$ ), Gaussian noise ( $\diamond$ ), Sharpening ( $\times$ ).

Table 9 displays the overall performance of the measures using two criteria: Correlation Coefficient (CC) and Root Mean Squared Error (RMSE) between MOS and objective prediction.

Table 9. Comparison of four measures

Criteria\Measure	PSNR	Q	MSSIM	M-DCT
CC	0.8038	0.8482	0.8282	0.8743
RMSE	7.0857	6.3096	6.6749	5.7821

The performance of the measures within each distortion type and across different distortion types are given in Tables 10 and 11, respectively.

We observe that the overall performance of M-DCT is better than the other three measures. For only Gaussian blur, the performance of M-DCT is slightly worse than the performances of Q and MSSIM. Across each distortion level, M-DCT outperforms all the other measures in comparison.

Table 10. (a) CC-based performance within each distortion type

Distortion type\Measure	PSNR	Q	MSSIM	M-DCT
JPEG	0.8877	0.9136	0.8768	0.9470
JPEG2000	0.7799	0.7810	0.8354	0.9156
Gaussian blur	0.8773	0.9124	0.9248	0.8868
Gaussian noise	0.9947	0.9585	0.9766	0.9955
Sharpening	0.9513	0.9662	0.9627	0.9931

(b) RMSE-based performance within each distortion type

Distortion type\Measure	PSNR	Q	MSSIM	M-DCT
JPEG	4.7647	4.2082	4.9762	3.3238
JPEG2000	3.5505	3.5387	3.1151	2.2783
Gaussian blur	6.1557	5.2497	4.8780	5.9259
Gaussian noise	0.9181	2.5556	1.9291	0.8485
Sharpening	1.0498	0.8783	0.9215	0.3984

Table 11. (a) CC-based performance across each distortion level

Distortion level\Measure	PSNR	Q	MSSIM	M-DCT
1	0.8024	0.7783	0.8436	0.9326
2	0.8246	0.8405	0.8542	0.9178
3	0.8361	0.8408	0.8402	0.8790
4	0.8422	0.8601	0.8549	0.8853
5	0.8358	0.8818	0.8755	0.8850

(b) RMSE-based performance across each distortion level

Distortion level\Measure	PSNR	Q	MSSIM	M-DCT
1	2.1151	2.2253	1.9028	1.2793
2	3.3381	3.1974	3.0679	2.3430
3	4.6552	4.5944	4.6026	4.0473
4	5.6156	5.3127	5.4026	4.8427
5	6.7775	5.8218	5.9646	5.7474

### 3. CONCLUSIONS

We presented three full-reference quality measures for color images. Each measure is based on a particular transform: (a) M-DFT uses the Discrete Fourier Transform, (b) M-DWT uses the Discrete Wavelet Transform, and (c) M-DCT uses the Discrete Cosine Transform. In all cases, the three measures, when the overall performance is considered, outperform the other measures (i.e., PSNR, Q, and MSSIM). The other experimental results can be summarized as follows:

- The performance of M-DFT is better than PSNR, Q, and MSSIM for each distortion type and distortion level.

- The performance of M-DWT is not as good as the performance of PSNR, Q, and MSSIM for three types of distortion (JPEG compression, JPEG compression, and Gaussian noise). However, the performance of M-DWT is considerably more consistent across distortion levels, providing higher CCs and lower RMSEs.
- The performance of M-DCT is slightly worse than the performances of Q and MSSIM for Gaussian blur. Across each distortion level, M-DCT outperforms all the other measures.

In future work, we will apply the proposed measures to watermarked images and video sequences.

## REFERENCES

1. A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 7-11, 2001.
2. Z. Wang, A. C. Bovik and B. L. Evans, "Blind measurement of blocking artifacts in images," *Proceedings of IEEE 2000 International Conferencing on Image Processing*, Vancouver, BC, Canada, September 10-13, 2000.
3. Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," *Proceedings of IEEE 2002 International Conferencing on Image Processing*, Rochester, NY, September 22-25, 2002.
4. L. Meesters and J.-B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Processing*, Vol. 82, pp. 369-387, 2002.
5. P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," *IEEE 2002 International Conference on Image Processing*, Rochester, NY, September 22-25, 2002.
6. E.-P. Ong, W. Lin, Lu, Z. Yang, S. Yao, F. Pan, L. Jiang and F. Moschetti, "A no-reference quality metric for measuring image blur," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
7. M. Carnec, P. Le Callet and D. Barba, "An image quality assessment method based on perception of structural information," *2003 International Conference on Image Processing*, Barcelona, Spain, September 14-17, 2003.
8. Z. Wang and E. P. Simoncelli, "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," *Human Vision and Electronic Imaging X Conference*, San Jose, CA, January 17-20, 2005, *Proceedings of SPIE*, Vol. 5666.
9. D. Van der Weken, M. Nachtegaal and E. E. Kerre, "A new similarity measure for image processing," *Journal of Computational Methods in Sciences and Engineering*, Vol. 3, No. 2, pp. 209-222, 2003.
10. A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
11. Z. Wang and A. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, Vol. 9, No. 3, pp. 81-84, March 2002.
12. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error measurement to structural similarity," *IEEE Transactions on Image Processing*, Vol. 13, No. 4, April 2004.
13. E. Girshtel, "A full-reference color image quality measure in the DFT domain," M.A. Dissertation, Department of CIS, CUNY Brooklyn College, 2900, Bedford Avenue, New York, NY 11210.
14. V. Slobodyan, "A full-reference color image quality measure in the DWT domain," M.A. Dissertation, Department of CIS, CUNY Brooklyn College, 2900, Bedford Avenue, New York, NY 11210.
15. J. Weissman, "A full-reference color image quality measure in the DCT domain," M.A. Dissertation, Department of CIS, CUNY Brooklyn College, 2900, Bedford Avenue, New York, NY 11210.
16. A. M. Eskicioglu and P. S. Fisher, "A Survey of Quality Measures for Gray Scale Image Compression," *Proceedings of 1993 Space and Earth Science Data Compression Workshop* (NASA Conference Publication 3191), pp. 49-61, Snowbird, UT, April 2, 1993.
17. A. M. Rohaly, J. Libert, P. Coriveau, and A. Webster (editors), "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," March 2000.

# Evaluating the Visual Quality of Watermarked Images

Aleksandr Shnayderman and Ahmet M. Eskicioglu\*

Department of Computer and Information Science

CUNY Brooklyn College, 2900 Bedford Avenue, Brooklyn, NY 11210

## ABSTRACT

A recent image quality measure, M-SVD, can express the quality of distorted images either numerically or graphically. Based on the Singular Value Decomposition (SVD), it consistently measures the distortion across different distortion types and within a given distortion type at different distortion levels. The SVD decomposes every real matrix into a product of three matrices  $A = USV^T$ , where  $U$  and  $V$  are orthogonal matrices,  $U^T U = I$ ,  $V^T V = I$  and  $S = \text{diag}(s_1, s_2, \dots)$ . The diagonal entries of  $S$  are called the singular values of  $A$ , the columns of  $U$  are called the left singular vectors of  $A$ , and the columns of  $V$  are called the right singular vectors of  $A$ . M-SVD, as a graphical measure, computes the distance between the singular values of the original image block and the singular values of the distorted image block, where  $n \times n$  is the block size. If the image size is  $k \times k$ , we have  $(k/n) \times (k/n)$  blocks. The set of distances, when displayed in a graph, represents a "distortion map." The numerical measure is derived from the graphical measure. It computes the global error expressed as a single numerical value. In this paper, we will extend the SVD-based image quality measure to evaluate the visual quality of watermarked images using several watermarking schemes.

Keywords: Singular Value Decomposition, image quality, watermarking, Discrete Wavelet Transform, Discrete Cosine Transform, Discrete Fourier Transform, Quantization Index Modulation, Lapped Orthogonal Transform.

## 1. INTRODUCTION

Measurement of image quality is a challenging problem in many image processing fields ranging from lossy compression to printing. The quality measures in the literature can be classified into two groups: Subjective and objective. Subjective evaluation is cumbersome as the human observers can be influenced by several critical factors such as the environmental conditions, motivation, and mood. The most common objective evaluation tool, the Mean Square Error (MSE), is very unreliable, resulting in poor correlation with the human visual system (HVS) [1]. In spite of their complicated algorithms, the more recent HVS-based objective measures do not appear to be superior to the simple pixel-based measures like the MSE, Peak Signal-to-Noise Ratio (PSNR), or Root Mean Squared Error (RMSE). It is argued that an ideal image quality measure should be able to describe [2]:

- (1) amount of distortion,
- (2) type of distortion, and
- (3) distribution of error.

Undoubtedly, there is a need for an objective measure that provides more information than a single numerical value. Only a few multi-dimensional measures exist in the relevant literature today [2].

Image quality measures can be classified using a number of criteria such as the type of domain (pixel or transform), the type of distortion predicted (noise, blur, etc.), and the type of information needed to assess the quality (original image, distorted image, etc.). Table 1 gives a classification based on these three criteria, and includes representative examples of recently published papers. Measures that require both the original image and the distorted image are called "full-reference" or "non-blind" methods, measures that do not require the original image are called "no-reference" or "blind" methods, and measures that require both the distorted image and partial information about the original image are called "reduced-reference" methods.

---

\* Corresponding author's email address: eskicioglu@sci.brooklyn.cuny.edu

We have recently developed a new measure, M-SVD, that can express the quality of distorted images either numerically or graphically [3,4,5]. Based on the Singular Value Decomposition (SVD), it consistently measures the distortion across different distortion types and within a given distortion type at different distortion levels. Comparison with the state-of-the-art metrics Q and MSSIM shows that the performance of M-SVD is superior.

Every real matrix  $A$  can be decomposed into a product of 3 matrices  $A = USV^T$ , where  $U$  and  $V$  are orthogonal matrices,  $U^T U = I$ ,  $V^T V = I$  and  $S = \text{diag}(s_1, s_2, \dots)$ . The diagonal entries of  $S$  are called the singular values of  $A$ , the columns of  $U$  are called the left singular vectors of  $A$ , and the columns of  $V$  are called the right singular vectors of  $A$ . This decomposition is known as the Singular Value Decomposition of  $A$ . It is one of the most useful tools of linear algebra with several applications to multimedia, including image compression and watermarking.

Table 1. Classification of image quality measures

Publication\Criterion	Domain type	Type of distortion predicted	Type of information needed
Van der Weken, Nachtegaal and Kerre, 2003 [6]	Pixel	Salt and pepper noise, enlightening and darkening	Full-reference
Beghdadi and Pesquet-Popescu, 2003 [7]	Discrete Wavelet Transform	Gaussian noise, grid pattern, JPEG compression	Full-reference
Bovik and Liu, 2001 [8]	Discrete Cosine Transform	JPEG compression	No-reference
Wang, Bovik and Evans, 2000 [9]	Discrete Fourier Transform	JPEG compression	No-reference
Wang, and Bovik, 2002 [10]	Pixel	JPEG compression	No-reference
Marziliano, Dufaux, Winkler and Ebrahimi, 2002 [11]	Pixel	Gaussian blur, JPEG 2000 compression	No-reference
Ong, Lin, Lu, Yang, Yao, Pan, Jiang and Moschetti, 2003 [12]	Pixel	Gaussian blur, JPEG 2000 compression	No-reference
Meesters and Martens, 2002 [13]	Pixel	JPEG compression	No-reference
Carnec, Le Callet and Barba, 2003 [14]	Pixel	JPEG compression, JPEG 2000 compression	Reduced-reference

M-SVD is a bivariate measure that computes the distance between the singular values of the original image block and the singular values of the distorted image block:

$$D_i = \text{Sqrt}[\sum_{i=1}^n (s_i - \hat{s}_i)^2],$$

where  $s_i$  are the singular values of the original block,  $\hat{s}_i$  are the singular values of the distorted block, and  $n \times n$  is the block size. If the image size is  $k \times k$ , we have  $(k/n) \times (k/n)$  blocks. The set of distances, when displayed in a graph, represents a "distortion map."

The numerical measure is derived from the graphical measure. It computes the global error expressed as a single numerical value, depending on the distortion type:

$$\text{M-SVD} = \frac{\sum_{i=1}^{(k/n) \times (k/n)} |D_i - D_{mid}|}{(k/n) \times (k/n)},$$

where  $D_{mid}$  represents the mid point of the sorted  $D_i$ s.



In this paper, we will extend the SVD-based image quality measure to evaluate the visual quality of watermarked images. We have developed three watermarking algorithms:

- A non-blind scheme using DWT and SVD [15]
- A non-blind scheme using DWT [16]
- A semi-blind scheme using DFT [17]

We also have access to the benchmarking suite [18] in the School of Electrical and Computer Engineering at Purdue University. On this web site, there are several watermarking schemes including:

- A non-blind scheme using DCT [19]
- A semi-blind scheme using DWT [20]
- A blind scheme based on the Quantization Index Modulation Algorithm (QIM) [21]
- A non-blind scheme using the Lapped Orthogonal Transform (LOT) based adaptive algorithm [22]

We will use M-SVD to compare the visual quality of watermarked images obtained by the above algorithms.

The embedding and detection algorithms for the watermarking schemes are as follows:

#### Reference 15:

##### Watermark embedding

1. Using DWT, decompose the cover image  $A$  into 4 subbands: LL, HL, LH, and HH.
2. Apply SVD to each subband image:  $A^k = U_a^k \Sigma_a^k V_a^{kT}$ ,  $k = 1,2,3,4$ , where  $k$  denotes LL, HL, LH, and HH bands, and  $\lambda_i^k$ ,  $i=1,\dots,n$  are the singular values of  $\Sigma_a^k$ .
3. Apply SVD to the visual watermark:  $W = U_w \Sigma_w V_w^T$ , where  $\lambda_{wi}$ ,  $i = 1,\dots,n$  are the singular values of  $\Sigma_w$ .
4. Modify the singular values of the cover image in each subband with the singular values of the visual watermark:  $\lambda_i^{*k} = \lambda_i^k + \alpha_k \lambda_{wi}$ ,  $i = 1,\dots,n$ , and  $k = 1,2,3,4$ .
5. Obtain the 4 sets of modified DWT coefficients:  $A^{*k} = U_a^k \Sigma_a^{*k} V_a^{kT}$ ,  $k = 1,2,3,4$ .
6. Apply the inverse DWT using the 4 sets of modified DWT coefficients to produce the watermarked cover image  $A'$ .

##### Watermark extraction

1. Using DWT, decompose the watermarked (and possibly attacked) cover image  $A^*$  into 4 subbands: LL, HL, LH, and HH.
2. Apply SVD to each subband image:  $A^{*k} = U_a^k \Sigma_a^{*k} V_a^{kT}$ ,  $k = 1,2,3,4$ , where  $k$  denotes the attacked LL, HL, LH, and HH bands.
3. Extract the singular values from each subband:  $\lambda_{wi}^k = (\lambda_i^{*k} - \lambda_i^k) / \alpha_k$ ,  $i = 1,\dots,n$ , and  $k = 1,2,3,4$ .
4. Construct the four visual watermarks using the singular vectors:  $W^k = U_w \Sigma_w^k V_w^T$ ,  $k = 1,2,3,4$ .

#### Reference 16:

##### Watermark embedding (first level decomposition)

1. Using two-dimensional separable dyadic DWT, obtain the first level decomposition of the cover image  $I$ .
2. Modify the DWT coefficients  $V_{ij}$  in the LL, HL, LH, and HH bands:  $V_{w,ij}^k = V_{ij}^k + \alpha_k W_{ij}$ ,  $ij=1, \dots, n$ , and  $k=1,2,3,4$ .
3. Apply inverse DWT to obtain the watermarked cover image  $I_w$ .

##### Watermark extraction (first level decomposition)

1. Using two-dimensional separable dyadic DWT, obtain the first level decomposition of the watermarked (and possibly attacked) cover image  $I_w^*$ .
2. Extract the binary visual watermark from the LL, HL, LH, and HH bands:  $W_{ij}^* = (V_{w,ij}^{*k} - V_{ij}^k) / \alpha_k$ ,  $ij=1, \dots, n$ , and  $k=1,2,3,4$ .
3. If  $W_{ij}^* > 0.5$ , then  $W_{ij}^* = 1$  else  $W_{ij}^* = 0$

##### Watermark embedding (second level decomposition)

1. Using two-dimensional separable dyadic DWT, obtain the second level decomposition of the cover image  $I$ .
2. Modify the DWT coefficients  $V_{ij}$  in the LL2, HL2, LH2, and HH2 bands:  $V_{w,ij}^k = V_{ij}^k + \alpha_k W_{ij}$ ,  $ij=1, \dots, n$ , and  $k=1,2,3,4$ .
3. Apply inverse DWT to obtain the watermarked cover image  $I_w$ .

##### Watermark extraction (second level decomposition)

1. Using two-dimensional separable dyadic DWT, obtain the second level decomposition of the watermarked (and possibly attacked) cover image  $I_w^*$ .
2. Extract the binary visual watermark from the LL2, HL2, LH2, and HH2 bands:  $W_{ij}^* = (V_{w,ij}^{*k} - V_{ij}^k) / \alpha_k$ ,  $ij=1, \dots, n$ , and  $k=1,2,3,4$ .
3. If  $W_{ij}^* > 0.5$ , then  $W_{ij}^* = 1$  else  $W_{ij}^* = 0$

#### Reference 17:

##### Watermark embedding

1. Convert the  $N \times N$  RGB cover image to YUV.
2. Compute the DFT of the luminance layer.
3. Move the origin to the center.
4. Obtain the magnitudes of DFT coefficients.
5. Divide the  $N \times N$  matrix of magnitudes into four  $(N/2) \times (N/2)$  matrices  $M_{ul}$ ,  $M_{ur}$ ,  $M_{ll}$ ,  $M_{lr}$ .  $ul$ : upper left,  $ur$ : upper right,  $ll$ : lower left,  $lr$ : lower right.
6. Define three frequency bands: low, middle, and high.
7. Embed a visual binary watermark in these three bands by determining the embedding locations.

8. In each band, repeat the following steps for all pairs of magnitudes:
  - a. Get a pair of magnitudes  $a$  and  $b$ :  $a$  from matrix  $M_{ul}$ , and the corresponding magnitude  $b$  from matrix  $M_{ur}$ .
  - b. Compute the mean  $m = (a+b)/2$ , and choose the value of the parameter  $p$ .
  - c. Embedding bit 1:
    - If  $a < m - (p/2 * m)$  then do not modify  $a$  and  $b$
    - else  $a = m - (p/2 * m)$  and  $b = m + (p/2 * m)$
  - d. Embedding bit 0:
    - If  $a > m + (p/2 * m)$  then do not modify  $a$  and  $b$
    - else  $a = m + (p/2 * m)$  and  $b = m - (p/2 * m)$
9. Copy the modified magnitudes in matrix  $M_{ul}$  to matrix  $M_{lr}$ .
10. Copy the modified magnitudes in matrix  $M_{ur}$  to matrix  $M_{ll}$ .
11. Obtain the DFT coefficients of the entire image using the modified magnitudes.
12. Compute the inverse DFT to obtain the luminance layer of the watermarked cover image.
13. Convert the YUV watermarked cover image to RGB.

Watermark extraction:

1. Convert the  $N \times N$  watermarked (and possibly attacked) RGB cover image to YUV.
2. Compute the DFT of the luminance layer.
3. Move the origin to the center.
4. Obtain the magnitudes of DFT coefficients.
5. Divide the  $N \times N$  matrix of magnitudes into four  $(N/2) \times (N/2)$  matrices  $M_{ul}$ ,  $M_{ur}$ ,  $M_{ll}$ ,  $M_{lr}$ .
6. Use the three frequency bands and the embedding locations defined in the embedding process: low, middle, and high.
7. For each pair of magnitudes in each band, if  $a > b$  then bit=0 else bit=1.

#### Reference 19:

Watermark embedding

1. Compute the DCT of the  $N \times N$  gray scale cover image  $I$ .
2. Embed a sequence of real values  $X = x_1, x_2, \dots, x_n$ , according to  $N(0,1)$ , into the  $n$  largest magnitude DCT coefficients, excluding the DC component:  $v_i' = v_i(1 + \alpha x_i)$ ,  $i = 1, \dots, n$ .
3. Compute the inverse DCT to obtain the watermarked cover image  $I'$ .

Watermark detection

1. Compute the DCT of the watermarked (and possibly attacked) cover image  $I^*$ .
2. Extract the watermark  $X^*$ :  $x_i^* = (v_i^* - v_i) / \alpha v_i$ ,  $i = 1, \dots, n$ .
3. Evaluate the similarity of  $X^*$  and  $X$  using  $\text{sim}(X, X^*) = \frac{X \cdot X^*}{(X^* \cdot X^*)^{1/2}}$ .
4. If  $\text{sim}(X, X^*) > T$ , a given threshold, the watermark exists.

#### Reference 20:

##### Watermark embedding

1. Compute the DWT of the  $N \times N$  gray scale cover image  $I$ .
2. Exclude the low pass DWT coefficients.
3. Embed the watermark into the DWT coefficients  $t_i > T_1$ :  $t'_i = t_i + \alpha |t_i| x_i$ , where  $i$  runs over all DWT coefficients  $> T_1$ .
4. Replace  $T = \{t_i\}$  with  $T' = \{t'_i\}$  in the DWT domain.
5. Compute the inverse DWT to obtain the watermarked cover image  $I'$ .

##### Watermark detection

1. Compute the DWT of the watermarked (and possibly attacked) cover image  $I^*$ .
2. Exclude the low pass DWT coefficients.
3. Select all the DWT coefficients higher than  $T_2$ .
4. Compute the sum  $z = \frac{1}{M} \sum_{i=1}^M y_i t_i^*$ , where  $i$  runs over all DWT coefficients  $> T_2$ ,  $y_i$  represents either the real watermark or a fake watermark,  $t_i^*$  represents the watermarked and possibly attacked DCT coefficients.
5. Choose a predefined threshold  $T_z = \frac{\alpha}{2M} \sum_{i=1}^M |t_i^*|$ .
6. If  $z$  exceeds  $T_z$ , the conclusion is the watermark is present.

#### Reference 21:

##### Watermark embedding

The encoder embeds the watermark into the host signal:  $S(x;m) = q\Delta(x + d(m)) - d(m)$ , where  $x$  is the host signal,  $q\Delta(x)$  is a uniform quantizer with quantization interval  $\Delta$ ,  $d(m)$  is the dither signal,  $m=0,1$  is the message bit. The dither signal  $d(0)$  and  $d(1)$  are constructed with the constraint

$$\begin{aligned} d(1) &= d(0) + \Delta/2 & \text{if } d[k,0] < 0, \\ d(1) &= d(0) - \Delta/2 & \text{if } d[k,0] \geq 0. \end{aligned}$$

This constraint ensures that the two dithered signals  $S(x;0)$  and  $S(x;1)$  have the maximum distance from each other.  $d(0)$  is the output from a uniform pseudo random number generator with range  $[-\Delta/2, \Delta/2]$ .

##### Watermark detection

The decoder decodes the watermark message  $m$  from the received composite signal  $S(x;m)$  without having access to the original host signal  $x$ . The message  $m$  that should be transmitted is the index for the quantizer used for quantizing the host-signal vector. While retrieving the hidden information, one evaluates a distance metric to all quantizers. The index of the quantizer with the smallest distance contributes to the message  $m$ .

## Reference 22:

### Watermark embedding

1. The Lapped Orthogonal Transform (LOT) divides the  $N \times N$  cover image  $I$  into overlapping  $16 \times 16$  blocks, and maps each block into an  $8 \times 8$  block in the frequency domain.
2. Introduce a perceptual analysis module to extract from each block a feature known as the Texture Masking Energy (TME). The blocks are classified into one of the following four categories according to the TME: texture, fine-texture, edge, and flat region. The edge blocks are further categorized according to the direction of the edges. The watermark embedding energy in each block is adjusted accordingly to adapt to the sensitivities of the HVS.
3. The quantization matrix for the  $k$ th block is  $Q_L(k) = Q_S \times M_{block}(k) \times M_{edge}(k)$ , where "x" denotes the element-wise multiplication.  $Q_S$  is the standard quantization matrix used in JPEG, and  $M_{block}(k)$ ,  $M_{edge}(k)$  are used to adjust the quantization steps in  $Q_S$  for the  $k$ th block.
4. In each block, the 5 most visually important AC coefficients bear the watermark:

$$X'_k(i_{n,k}, j_{n,k}) = X_k(i_{n,k}, j_{n,k}) + \alpha Q_L(i_{n,k}, j_{n,k}) w(n),$$

where  $X$  is the LOT coefficient of the original image,  $X'$  is the corresponding watermarked LOT coefficient,  $w(n)$  is the watermark element, and  $(i_{n,k}, j_{n,k})$  denotes the specified position to bear the watermark in the  $k$ th block.

5. Compute the inverse LOT to obtain the watermarked cover image  $I'$ .

### Watermark detection

1. Extract the watermark using the LOT coefficients of the watermarked (and possibly attacked) cover image  $I^*$  and the LOT coefficients of the original cover image  $I$ :  $\hat{w}(n) = (\hat{X}'_k(i_{n,k}, j_{n,k}) - X_k(i_{n,k}, j_{n,k})) / (\alpha Q_L(i_{n,k}, j_{n,k}))$ .
1. Compare the extracted watermark  $\hat{W}$  and the original watermark  $W$  using the formula  $sim(W, \hat{W}) = \frac{W \cdot \hat{W}}{\sqrt{W \cdot W}}$ .
2. If the value of  $sim(W, \hat{W})$  is larger than a specified threshold, the watermark is present.

## 2. EXPERIMENTS

The original  $512 \times 512$  images used with the watermarking algorithms are shown in Figure 1.

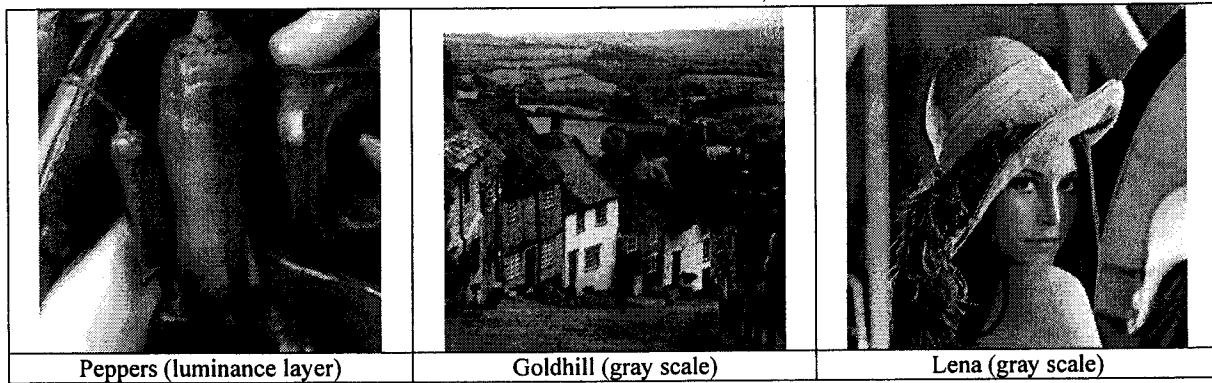


Figure 1. Original cover images

The luminance layer of each full color watermarked image, the watermarked gray scale images, the corresponding distortion maps, and the values of the numerical measure are given in Figure 2.


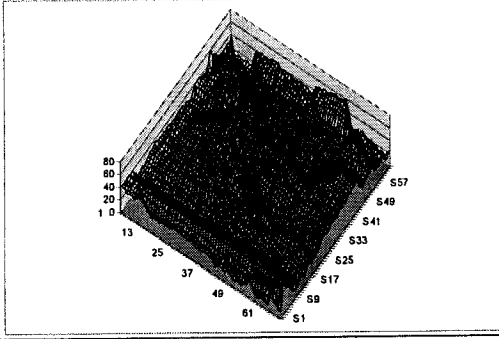

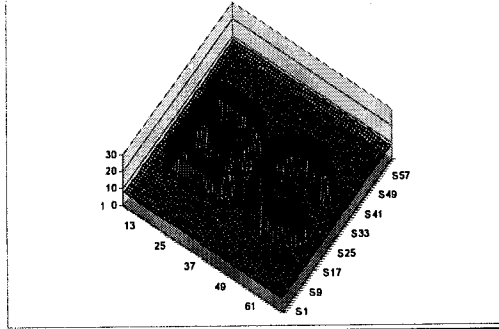

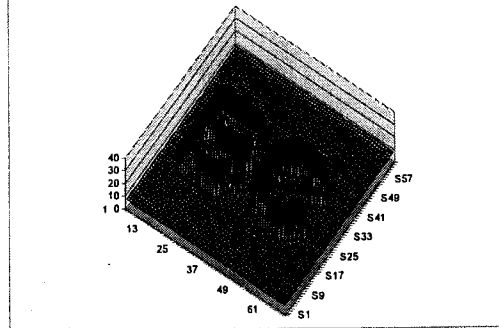

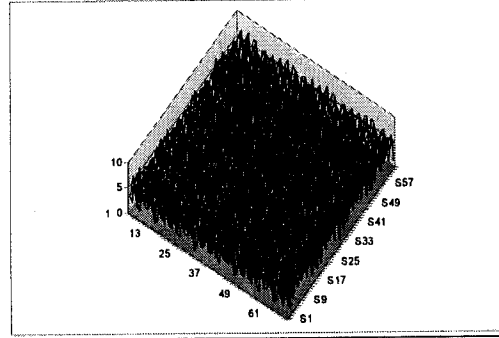

Watermarked image	Distortion map and global error
	 <p>M-SVD = 9.684</p>
<p>Reference 15</p> 	 <p>M-SVD = 3.095</p>
<p>Reference 16 (first level)</p> 	 <p>M-SVD = 3.313</p>
<p>Reference 16 (second level)</p> 	 <p>M-SVD = 1.298</p>
<p>Reference 17</p> 	

Figure 2. Watermarked images, distortion maps, and numerical measure values


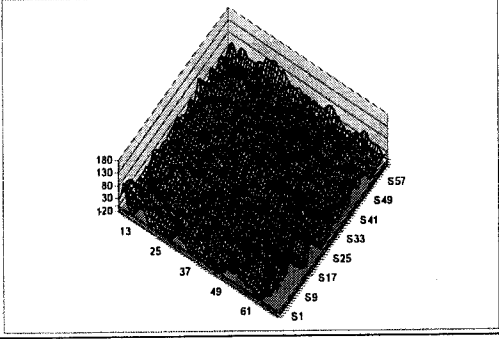

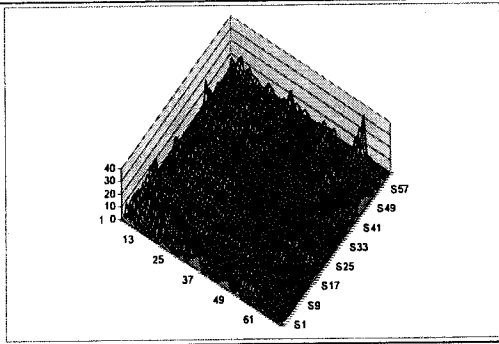

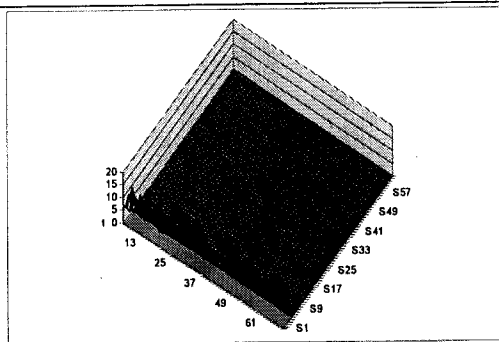
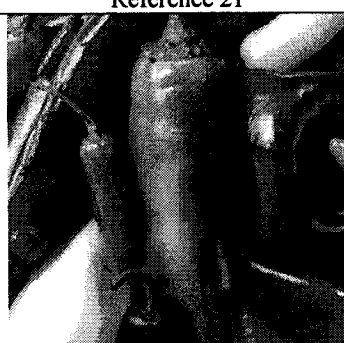
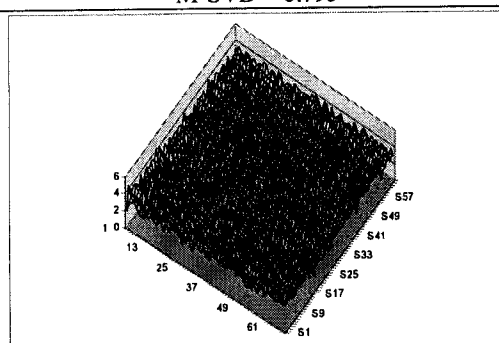

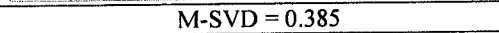
Watermarked image	Distortion map and global error
	 <p>M-SVD = 16.152</p>
Reference 19 	 <p>M-SVD = 16.152</p>
Reference 20 	 <p>M-SVD = 1.615</p>
Reference 21 	 <p>M-SVD = 0.795</p>
Reference 22 	 <p>M-SVD = 0.385</p>

Figure 2. Continued

For the Quantization Index Modulation (QIM) algorithm [21], the text embedded in the color image Peppers is:

Brooklyn College  
The City University of New York  
2900 Bedford Avenue  
Brooklyn, NY 11210

### 3. CONCLUSIONS

A recent measure, M-SVD, can express the quality of distorted images either numerically or graphically. Based on the Singular Value Decomposition (SVD), it consistently measures the distortion both across different distortion types and within a given distortion type at different distortion levels. The SVD decomposes every real matrix into a product of 3 matrices  $A = USV^T$ , where  $U$  and  $V$  are orthogonal matrices,  $U^T U = I$ ,  $V^T V = I$  and  $S = \text{diag}(s_1, s_2, \dots)$ . The diagonal entries of  $S$  are called the singular values of  $A$ , the columns of  $U$  are called the left singular vectors of  $A$ , and the columns of  $V$  are called the right singular vectors of  $A$ .

M-SVD, as a graphical measure, computes the distance between the singular values of the original image block and the singular values of the distorted image block, where  $n \times n$  is the block size. If the image size is  $k \times k$ , we have  $(k/n) \times (k/n)$  blocks. The set of distances, when displayed in a graph, represents a "distortion map." The numerical measure is derived from the graphical measure. It computes the global error expressed as a single numerical value.

In this paper, we extended the SVD-based image quality measure to evaluate the visual quality of watermarked images using seven algorithms.

Even if the hacker knows the algorithm for M-SVD, she will not be able to understand the algorithms for [17], [19], and [22] as these algorithms produce distortion maps that appear to be random.

For [15], [16], [20], and [21], the graphical measure indicates what has been embedded.

- For the DWT-SVD domain algorithm [15], the contours of Lena are apparent.
- For the DWT domain algorithm [16], the embedded binary logo ("BC") is definitely visible.
- For the DWT domain algorithm [20], which embeds the watermark in LH, HL, and HH bands, the edges of the peppers can be noticed because these three bands correspond to higher frequency areas.
- For QIM [21], if the embedded text is longer, the graphical measure adds more columns in the distortion map.

The numerical measure computes a global error with values that range from 0.385 to 16.152, depending on what is embedded in the cover images. The smallest and largest values are computed for [22] and [19], respectively, probably because of the following reasons:

- In [22], an HVS model is used, and only 5 most visually important AC coefficients are modified in each  $8 \times 8$  block in the LOT domain.
- In [19], no HVS model is used, and the highest 1000 AC coefficients are modified in the DCT domain.

In future work, we will apply M-SVD to video sequences such as akiyo, flowergarden, and tennis.

### ACKNOWLEDGEMENT

This work was supported in part by a grant from the Air Force Research Lab (AFRL) under Award FA9550-05-1-0400.



## REFERENCES

- [1] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Transactions on Communications*, Vol. 43, pp. 2959-2965, December 1995.
- [2] A. M. Eskicioglu, "Quality measurement for monochrome compressed images in the past 25 years," *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 4, pp. 1907-1910, Istanbul, Turkey, June 5-9, 2000.
- [3] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "A multidimensional image quality measure using singular value decomposition," *Proceedings of SPIE Image Quality and System Performance*, Vol. 5294, pp. 82-92, San Jose, CA, January 19-20, 2004.
- [4] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-based gray-scale image quality measure for local and global assessment," *IEEE Transactions on Image Processing*, 2006.
- [5] A. Shnayderman and A. M. Eskicioglu, "Assessment of full-color image quality with singular value decomposition," *IS&T/SPIE's 17th Symposium on Electronic Imaging, Image Quality and System Performance II Conference*, San Jose, CA, January 16-20, 2005.
- [6] D. Van der Weken, M. Nachtgael and E. E. Kerre, "A new similarity measure for image processing," *Journal of Computational Methods in Sciences and Engineering*, Vol. 3, No. 2, pp. 209-222, 2003.
- [7] A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
- [8] A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Salt Lake City, UT, May 7-11, 2001.
- [9] Z. Wang, A. C. Bovik and B. L. Evans, "Blind measurement of blocking artifacts in images," *Proceedings of IEEE 2000 International Conferencing on Image Processing*, Vancouver, BC, Canada, September 10-13, 2000.
- [10] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," *Proceedings of IEEE 2002 International Conferencing on Image Processing*, Rochester, NY, September 22-25, 2002.
- [11] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," *IEEE 2002 International Conference on Image Processing*, Rochester, NY, September 22-25, 2002.
- [12] E.-P. Ong, W. Lin, Lu, Z. Yang, S. Yao, F. Pan, L. Jiang and F. Moschetti, "A no-reference quality metric for measuring image blur," *7th International Symposium on Signal Processing and Its Applications*, Paris, France, July 1-4, 2003.
- [13] L. Meesters and J.-B. Martens, "A single-ended blockiness measure for JPEG-coded images," *Signal Processing*, Vol. 82, pp. 369-387, 2002.
- [14] M. Carnec, P. Le Callet and D. Barba, "An image quality assessment method based on perception of structural information," *2003 International Conference on Image Processing*, Barcelona, Spain, September 14-17, 2003.
- [15] E. Ganic and A. M. Eskicioglu, "Robust DWT-SVD domain image watermarking: embedding data in all frequencies," *ACM Multimedia and Security Workshop 2004*, Magdeburg, Germany, September 20-21, 2004.

- [16] P. Tao and A. M. Eskicioglu, "A robust multiple watermarking scheme in the DWT domain," *Optics East 2004 Symposium, Internet Multimedia Management Systems V Conference*, Philadelphia, PA, October 25-28, 2004.
- [17] J. Kusyk and A.M. Eskicioglu, "An effective blind watermarking scheme for color images through comparison of DFT coefficients," *Optics East 2005 Symposium, Internet Multimedia Management Systems VI Conference*, Boston, MA, October 23-26, 2005.
- [18] H. C. Kim, H. Ogunley, O. Guitart and E. J. Delp, "The Watermark Evaluation Testbed," *Proceedings of SPIE Electronic Imaging, Security, Steganography, and Watermarking of Multimedia Contents VI*, San Jose, CA, January 19-22, 2004.
- [19] I. J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, Vol. 6, No. 12, pp. 1673-1687, 1997.
- [20] R. Dugad, K. Ratakonda, and N. Ahuja, "A new wavelet-based scheme for watermarking images," *Proceedings of the International Conference on Image Processing*, October 4-7, 1998, Vol. 2, pp. 419-423.
- [21] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *Journal of VLSI Signal Processing*, Vol. 27, pp. 7-33, 2001.
- [22] Y. Liu, B. Ni, X. Feng, and E. J. Delp, "LOT-based adaptive image watermarking," *SPIE International Conference on Security, Steganography, and Watermarking of Multimedia Contents VI*, January 18-22, 2004, San Jose, CA.

# Video Quality Assessment Using M-SVD

Peining Tao<sup>a</sup>, Ahmet M. Eskicioglu<sup>b</sup>

<sup>a</sup>Dept. of Computer Science, Graduate Center, City University of New York,  
365 Fifth Avenue, New York, NY USA 10016-4309;

<sup>b</sup>Dept. of Computer Information Science, Brooklyn College, City University of New York,  
2900 Bedford Ave., Brooklyn, NY USA 11201-9642

## ABSTRACT

Objective video quality measurement is a challenging problem in a variety of video processing application ranging from lossy compression to printing. An ideal video quality measure should be able to mimic the human observer. We present a new video quality measure, M-SVD, to evaluate distorted video sequences based on singular value decomposition. A computationally efficient approach is developed for full-reference (FR) video quality assessment. This measure is tested on the Video Quality Experts Group (VQEG) phase I FR-TV test data set. Our experiments show the graphical measure displays the amount of distortion as well as the distribution of error in all frames of the video sequence while the numerical measure has a good correlation with perceived video quality outperforms PSNR and other objective measures by a clear margin.

**Keywords:** video quality measure, subjective evaluation, singular value decomposition, M-SVD, full-reference, Video Quality Experts Group(VQEG), PSNR

## 1. INTRODUCTION

Objective image and video quality measures play an important role in variety of image/video processing applications ranging from lossy compression to printing. Quality measures can be classified into two state-of-art category: Subjective and objective [1]. Subjective evaluation is cumbersome as the human observers can be influenced by several critical factors such as the environmental conditions, motivation and mood. Object is considerably stable but may not correlate well with Human Visual System [2,3,4,5,6,7]. Objective measures in the literature can be classified into three types according to the type of information needed during quality assessment: Measures that require both the original video and the distorted video are called "full-reference" or "non-blind" methods [8,9], measures that do not require the original video are called "no-reference" or "blind" methods [10,11,12,13,14], and measures that require both the distorted video and partial information about the original video are called "reduced-reference" methods [15]. Currently, the most commonly used full-referenced objective evaluation tools, the Mean Square Error (MSE) and Peak Signal-to-Noise Ratio (PSNR), are very unreliable, resulting in poor correlation with the HVS. Many efforts have been made to design image/video quality assessment models, incorporating perceptual quality measures by considering the characteristics of HVS. In spite of their complicated algorithms, the more recent HVS-based objective measures do not appear to be superior to the simple pixel-based MSE and PSNR. There is an increasing need to develop an objective quality measure that may predict the perceived video quality automatically. Moreover, it is argued that an ideal image/video quality measure should be able to describe the amount of distortion as well as the distribution of error. Undoubtedly, there is a need for an objective measure that provides more information than a single numerical value. Very few multi-dimensional measures exist in the relevant literature today.

A recent paper [16] developed a new measure, M-SVD, can express the quality of distorted images either numerically or graphically. Comparison with the state-of-the-art metrics UQI [17] and MSSIM [18] shows that the performance of M-SVD is superior. In this paper, we propose an improved version of the algorithm and employed for video quality assessment. The proposed algorithm is applied to the VQEG Phase I [19,20] test dataset for FR-TV video quality assessment. The Video Quality Experts Group (VQEG) was formed to develop, validate and standardize new objective measurement methods for video quality. The introduction to the new video quality assessment measure is given in Section 2. The measure is tested on the VQEG Phase I FR-TV video dataset. Experimental results are presented in Section 3. Finally, Section 4 provides the brief conclusion and future discussion.

## 2. QUALITY MEASURES USING M-SVD

### 2.1 M -SVD for gray scale images

A recently developed measure, M-SVD [16,21,22], that can express the quality of distorted images either numerically or graphically. Based on the Singular Value Decomposition (SVD) [23,24], it consistently measures the distortion both across different distortion types and within a given distortion type at different distortion levels. Comparison with the state-of-the-art metrics UQI and MSSIM shows that the performance of M-SVD is superior.

Every real matrix  $A$  can be decomposed into a product of 3 matrices  $A = USV^T$ , where  $U$  and  $V$  are orthogonal matrices,  $U^T U = I$ ,  $V^T V = I$  and  $S = \text{diag}(s_1, s_2, \dots)$ . The diagonal entries of  $S$  are called the singular values of  $A$ , the columns of  $U$  are called the left singular vectors of  $A$ , and the columns of  $V$  are called the right singular vectors of  $A$ . This decomposition is known as the Singular Value Decomposition of  $A$ . It is one of the most useful tools of linear algebra with several applications to multimedia including image compression and watermarking.

The proposed graphical measure is a bivariate measure that computes the distance between the singular values of the original image block and the singular values of the distorted image block:

$$D_k = \text{Sqrt}[\sum_{i=1}^n (s_i - \hat{s}_i)^2] \quad (1)$$

where  $s_i$  and  $\hat{s}_i$  are the singular values of the original block and distorted block,  $D_k$  is the distance of singular values of the  $k^{\text{th}}$  distorted block, and  $n$  is the block size. If the image size is  $rxn$ , we have  $(r/n) \times (c/n)$  blocks. The set of distances, when displayed in a graph, represents a "distortion map."

The numerical measure is derived from the graphical measure. It computes the global error expressed as a single numerical value depending on the distortion map:

$$M\text{-SVD} = \frac{\sum_{i=1}^{(k/n) \times (k/n)} |D_i - D_{mid}|}{(r/n) \times (c/n)} \quad (2)$$

where  $D_{mid}$  represents the mid point of the sorted  $D_i$ ,  $rxn$  is the image size, and  $n$  is the block size.

The measure is employed to 512x512 gray-scale images. The distortion types, the distortion levels and the associated parameters are listed in Table 1. Fig. 1 shows the distorted lena size of 512x512 across 6 different types with 5 different levels. Fig. 2 presents the distortion maps, which provides the amount of distortion, the type of distortion, and the distribution of error, are obtained as gray-scale images by mapping the  $D_i$  values to the range [0,255]. In their experiments, the authors choosed 8 as block size. Although the size of the distortion map is 64x64, they are enlarged to the size of distorted image to make pixel values more visible. The darker pixel values indicate small distances, and the lighter pixel values indicate larger differences. The global error expressed as a single numerical value (provides overall error based on the distortion) as presented in equation 2. Pearson correlation between the single numerical value with subjective evaluation is computed. Analysis shows the performance of M-SVD is much better than PSNR, UQI and MSSIM. Table2 and Table 3 display the correlation coefficients between subjective evaluation and M-SVD in comparison with other objective models across different distortion types in different distortion levels.

Table 1. Distortion types and levels applied to tested image.

Type \ Level	Level 1	Level 2	Level 3	Level 4	Level 5
JPEG	10:1	20:1	30:1	40:1	50:1
JPEG2000	10:1	20:1	30:1	40:1	50:1
Gaussian blur	1	2	3	4	5
Gaussian noise	3	6	9	12	15
Sharpening	10	20	30	40	50
DC-shifting	4	8	12	16	20

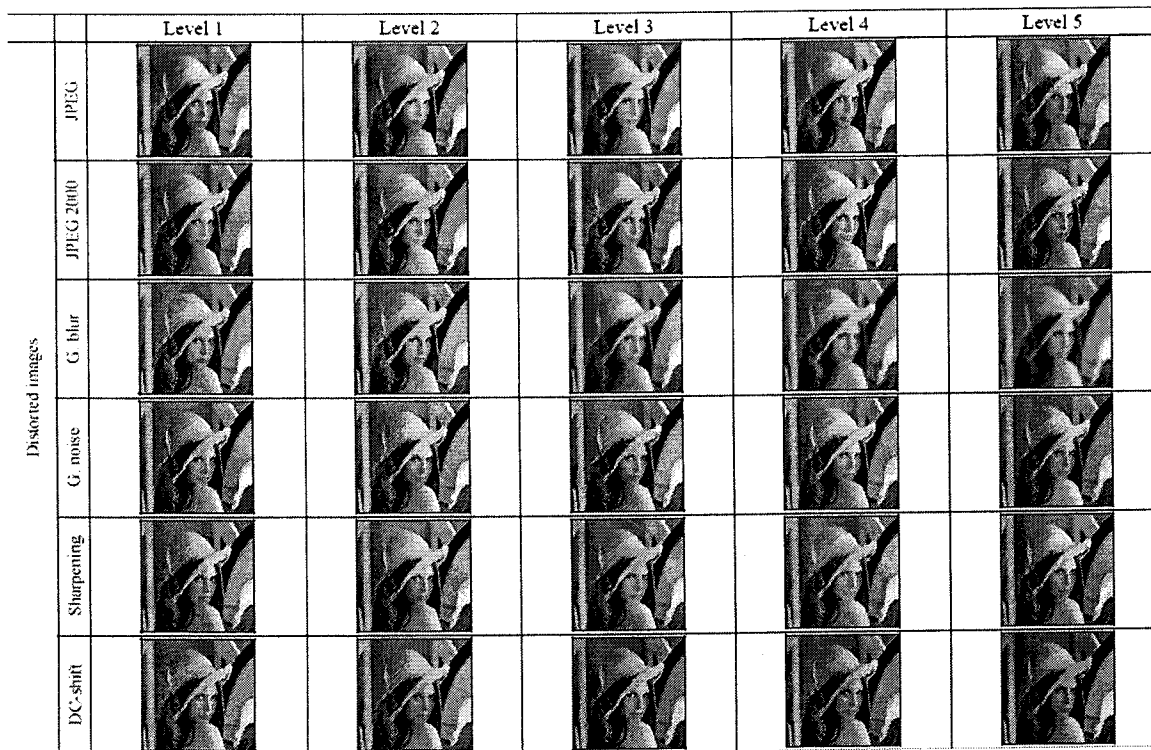


Fig. 1. The distorted lena across 6 different types in 5 different levels.

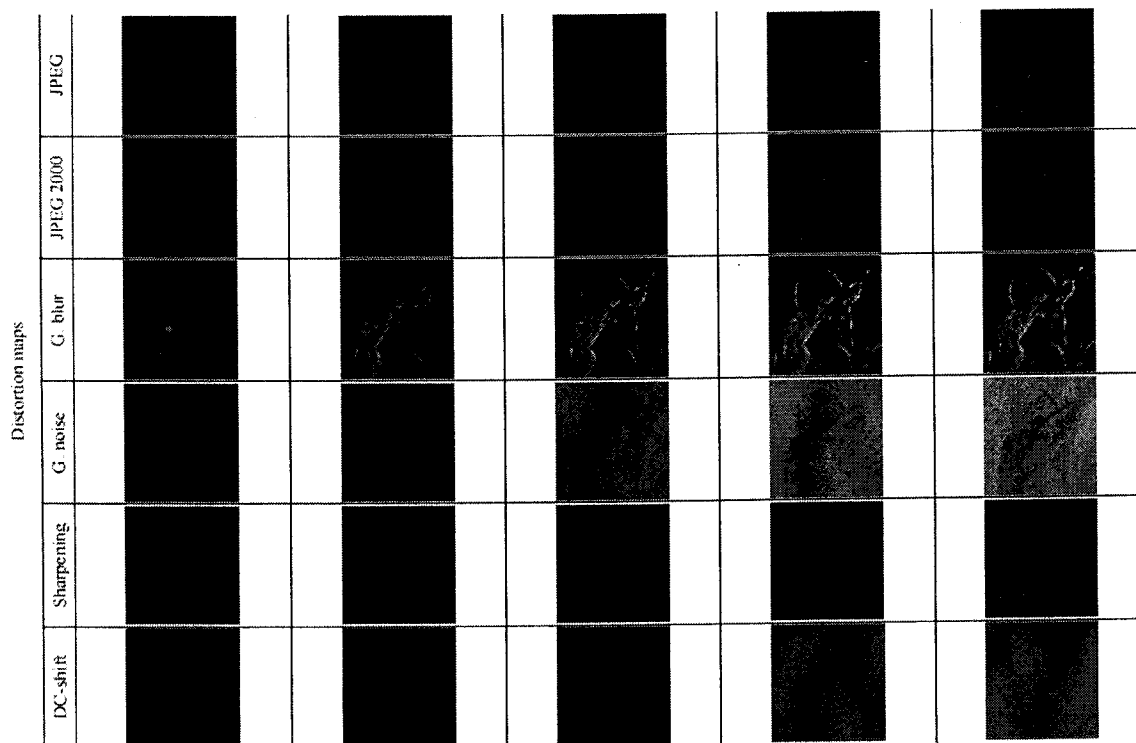


Fig. 2. The corresponding distortion maps of the distorted images in Fig1.

Table. 2. Correlation coefficients between subjective evaluation and M-SVD in comparison with other objective models across each distortion type.

Distortion Type\Measure	PSNR	UQI	MSSIM	M-SVD
JPEG	0.974	0.904	0.928	0.977
JPEG2000	0.949	0.688	0.801	0.952
Gaussian blur	0.816	0.917	0.906	0.929
Gaussian noise	0.901	0.984	0.987	0.975
Sharpening	0.955	0.908	0.947	0.937
DC-shifting	0.914	0.637	0.643	0.718

Table. 3. Correlation coefficients between subjective evaluation and M-SVD in comparison with other objective models across each distortion level.

Distortion Type\Measure	PSNR	UQI	MSSIM	M-SVD
1	0.808	0.744	0.781	0.890
2	0.751	0.808	0.853	0.954
3	0.529	0.885	0.910	0.962
4	0.369	0.914	0.929	0.958
5	0.439	0.940	0.947	0.924

## 2.2 Video Quality Assessment Algorithm

The framework of video quality assessment system is shown in Fig.4. First, both the original and processed videos are subjected to pre-processing procedure. The original and processes video sequences are resample to 4:4:4, Y, Cb, Cr format. A spatial-temporal-luminance alignment is included into the system to normalize the input sequences. In the second step, a set of frames of a video sequence are divided into 8x8 blocks. The block-based SVD is computed on the luminance layer (Y) [10] in both the original frames and the corresponding processed frames. Then, the local error measure as equation 1 computes the distance between the singular values of the original frame block and the singular values of the distorted frame block. It is well known that human eyes are over-sensitive to high contrast areas, especially for edges. The edges are computed by calculating the local gradient of the luminance signal (using a Sobel like spatial filtering) in the luminous layer. An edge function is used to detect edges, which returns a binary image containing 1's

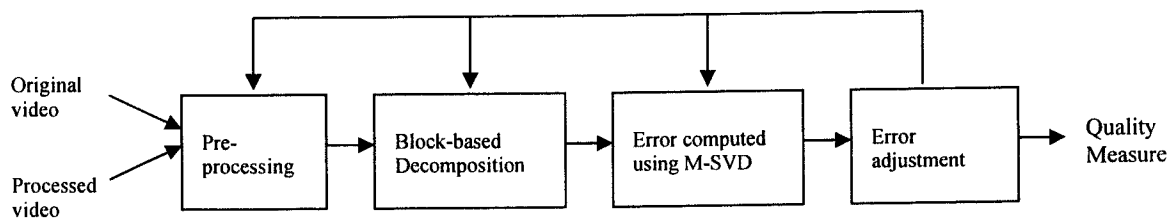


Fig. 4. The frame work of video quality assessment system

where edges are found and 0's elsewhere. This output binary image allows us to assign an edge index to each block within a frame. Those blocks in a frame correspond to object boundaries as edges receive higher edge index as follows:

$$B_k = \frac{\frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n b_{ij}}{\frac{1}{r \times c} \sum_{i=1}^r \sum_{j=1}^c b_{ij}} \quad (3)$$

and

$$B_k = \begin{cases} B_k & \text{if } B_k \geq \tau \\ 0 & \text{if } B_k < \tau \end{cases} \quad (4)$$

where  $B_k$  denotes the edge index of  $k^{th}$  block,  $r \times c$  denotes the frame dimension,  $n$  is the block size, and  $\tau$  is a threshold that determines the strength of the edge index. The local error measure with edge detection is given by:

$$D_k = B_k \text{Sqrt}[\sum_{i=1}^n (s_i - \hat{s}_i)^2] \quad (5)$$

The frame-level error measure described in equation 2 is expressed as one single value depending on local error measures. This measure is applied to  $Y$ ,  $Cb$ ,  $Cr$  components respectively and combined to obtain the frame-level error measure using a weighted summation. Let  $M-SVD_j^Y$ ,  $M-SVD_j^{Cb}$  and  $M-SVD_j^{Cr}$  be the error measures of  $Y$ ,  $Cb$ ,  $Cr$  components of the  $j^{th}$  frame, respectively. The combined quality error index is:

$$M-SVD_j = W_Y M-SVD_j^Y + W_{Cb} M-SVD_j^{Cb} + W_{Cr} M-SVD_j^{Cr} \quad (6)$$

where the weights  $W_Y$ ,  $W_{Cb}$  and  $W_{Cr}$  are obtained experimentally. In the above formula, the contributions from the luminance and two chrominance components are, respectively, 80%, 10%, and 10%, Therefore, the luminous component makes the major contribution.

The local error output can be mapped to a gray scale image as values in the range [0, 255]. Therefore, the local error is expressed as a 3-dimensional graph that provides the amount of the error as well as its distribution in a frame. Such graph results in a detailed distortion map leading to higher correlation with subjective evaluation. The frame-level error is expressed as a single numerical value. A plot of the error series of all frames contained in video sequence graphically describes the amount and the distribution of error in the distorted sequence.

Finally the overall quality of the entire sequence is defined as:

$$M-SVD = \frac{\sum_{i=1}^n T_i M-SVD_i}{n} \quad (7)$$

where  $n$  is the number of frames in a video sequence.  $T_i$  is the weighting value assigned to  $j^{th}$  frame in a video sequence. This leads to a quality measure that is equal to the weighted average  $M-SVD$  error measure of all frames. Error adjustment is incorporated throughout the all steps in the video quality assessment system. Since the HVS is a complicated system, error adjustment is attempting to improve the performance of the quality assessment system by incorporating perceived quality features. In our proposed system, error adjustment is expressed in the edge index and weighting factors applied to both local and global error assessment as well as the sequence calibration in the pre-processing step.

### 3. EXPERIMENTAL RESULTS

The proposed algorithm is applied to the VQEG Phase I test dataset for FR-TV video quality assessment. The Video Quality Experts Group (VQEG) was formed to develop, validate and standardize new objective measurement methods for video quality. They collected a large set of subjective-rated video data where over 26,000 subjective opinion scores were generated based on 20 different source sequences processed by 16 different video systems and evaluated at eight independent laboratories worldwide. VQEG provides this valuable dataset, which we can utilize to evaluate and improve object video quality measurement.

### 3.1 Graphical measure

The original and processed video sequences are resampled to 4:4:4, Y, Cb, Cr format. A spatial-temporal-luminance alignment is applied in the system to normalize the input sequences. In our experiment, all frames of the input video sequence are divided into 8x8 blocks. We computed SVD in all blocks of the luminance and color components respectively in both the original frames and the corresponding processed frames. The local error measure specified in equation 1 computes the distance between the singular values of the original frame block and the singular values of the distorted frame block. The local error output can be mapped to a gray scale image as values in the range [0, 255]. Fig 5 shows the distortion maps as 2-dimensional and 3-dimensional graphs that provides the amount of the error as well as its

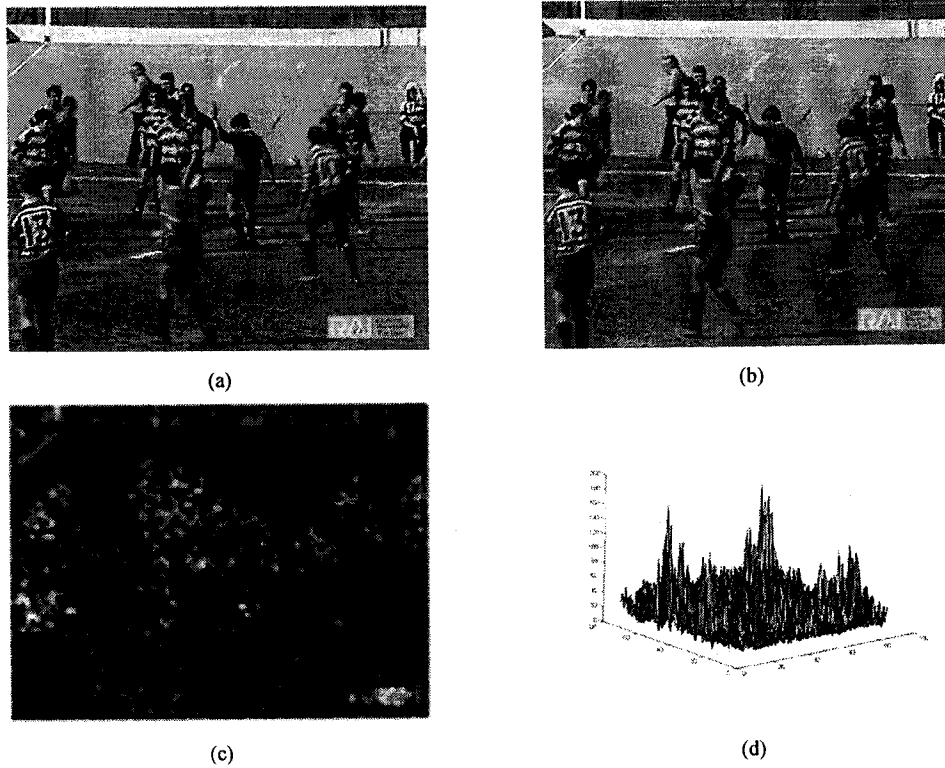


Fig. 5. The distortion maps as a 2 and 3-dimensional graphs for one frame in luminous layer. (a) original frame size of 568x680 (b) processed frame size of 568x680 (c) 2-dimensional distortion map size of 71x85 (d) 3-dimensional distortion map.

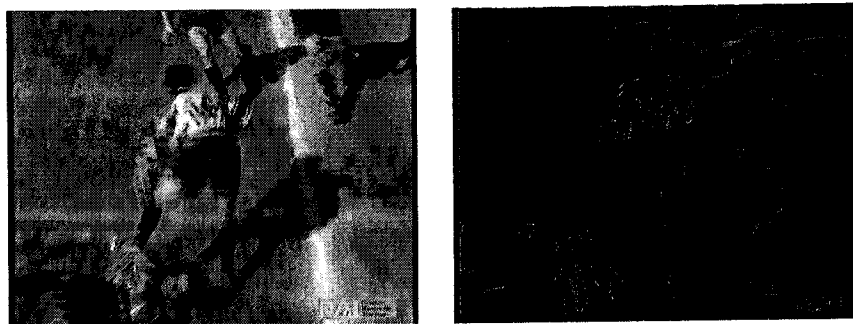


Fig. 6. The binary image on the right as the output of the edge function of one frame on the left.



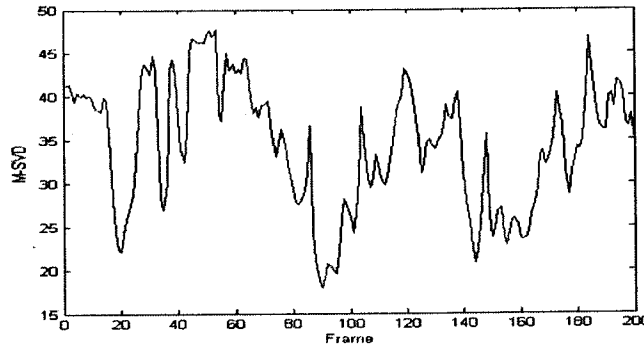


Fig. 7. The error series of all frames contained in one distorted video sequence.

distribution in a frame. It is known that naked eyes are highly alert to the distortion on edges. An edge function is applied by calculating the local gradient of the luminance signal (using a Sobel like spatial filtering) in the luminous layer. Fig 6 displays the binary image as the output of the edge function applied one frame in a video sequence. Each block inside a frame is received an edge index as specified in equations 3, 4 and 5. The frame-level error measure is computed as equation 2 for Y, Cb and Cr components respectively depending upon local error measures adjusted with edge index. The outputs are combined to obtain the overall frame-level error using a weighted summation. Fig 7 shows the graphical error measure of one distorted video sequence. In the graph, the horizontal axis is the frame index of one video sequence; the vertical axis is M-SVD error measure output of the frames in such sequence. Therefore, it provides the amount and distribution of error through the distorted sequence.

### 3.2 Numerical measure

The overall quality of video sequence is expressed as one single numerical value using equation 7. We follow the procedures employed in the VQEG Phase I objective video quality model test plan to evaluate the performance of our system. Four metrics are used in the evaluation of objective results. Metric 1 is the correlation coefficient between objective and subjective scores after variance-weighted regression analysis, including a test of significance of the difference. Metric 2 is the correlation coefficient between objective and subjective scores after non-linear regression analysis. A logistic function used in the first two metrics provides a non-linear map between subjective and objective scores, where a weighted least squares procedure was applied to the logistic function. The first two metrics predict the accuracy of an objective model. Metric 3 is the Spearman rank-order correlation coefficient between the objective and subjective scores. This correlation method only assumes a monotonic relationship between the two scores. Metric 3 is related to prediction monotonicity of a model. Metric 4 is outlier ratio of "outlier-points" to total number of points. The model's prediction consistency can be measured by the number of outlier points (defined as having an error greater than some threshold as a fraction of the total number of points). A smaller outlier fraction means the model's predictions are more consistent.

Table I presents the performance comparison of video quality assessment models on VQEG Phase I Test Data Set (all test video sequences included). P1-P9 [19] are nine different proponent models submitted to VQEG for evaluation. P0 (PSNR) is included by VQEG as a reference objective model. SSIM (Structural Similarity Index) [25] is a new measure presented in a recent paper. Model M-SVD/Y computes the difference between the original and the distorted frames on the luminance (Y) component using equations 1, 2, and 7. Model M-SVD/YCbCr is applied to Y, Cb, Cr color components independently. The frame level error measures for each component are combined using equation 6. Model M-SVD/Edge Detection is applied to Y, Cb and Cr with edge detection performed on Y component using equations 2, 5, 6 and 7. Table I displays the comparison results of the four metrics for different models. It can be observed that the proposed M-SVD measure outperforms all other measures by a clear margin. Figure 7 gives the non-linear regression analysis of the subjective/objective scores on all video sequences in the VQEG Phase I test given by PSNR, M-SVD/Y, M-SVD/YCbCr and M-SVD/Edge Detection. Each graph shows the scatter plot of subjective and objective scores and the fitted curve. 160 video sequences are tested for each objective model. In each graph associated with one model,

Table. 2. Performance comparison of video quality assessment models on VQEG Phase I Test Data Set (all test video sequences included).

Model	Metric 1	Metric 2	Metric 3	Metric 4
P0 (PSNR)	0.804	0.779	0.786	0.678
P1 (CPqD)	0.777	0.794	0.781	0.650
P2 (T/S)	0.792	0.805	0.792	0.656
P3 (NHK)	0.726	0.751	0.718	0.725
P4 (KDD)	0.622	0.624	0.645	0.703
P5 (EPFL)	0.778	0.777	0.784	0.611
P6 (TAPESTRIES)	0.277	0.310	0.248	0.844
P7 (NASA)	0.792	0.770	0.786	0.636
P8 (KPN)	0.845	0.827	0.803	0.578
P9 (NTIA)	0.781	0.782	0.775	0.711
SSIM	0.864	0.849	0.812	0.578
M-SVD/Y Component	0.891	0.868	0.752	0.531
M-SVD/ YCbCr	0.891	0.871	0.766	0.514
M-SVD/Edge Detection	0.893	0.877	0.799	0.486

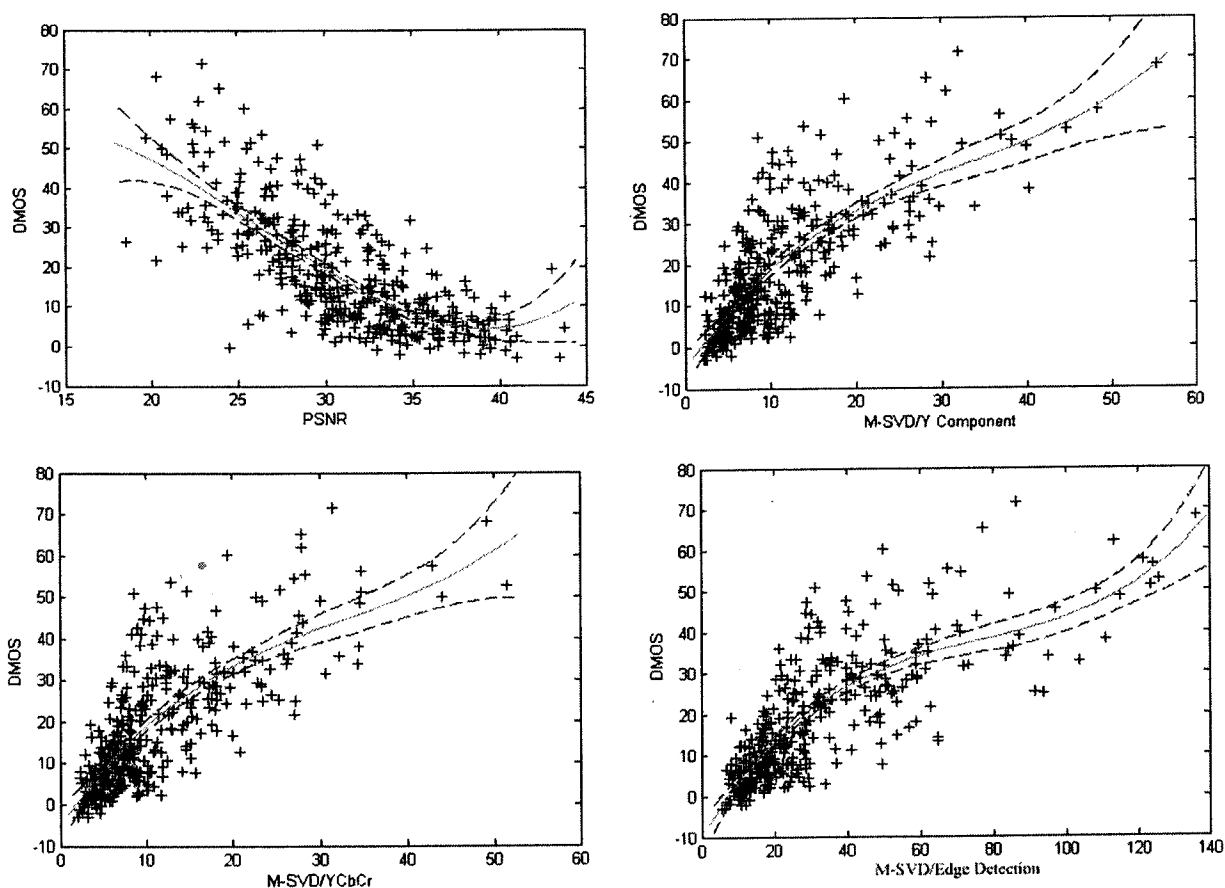


Fig. 7. The Scatter plot comparison of objective models on all video sequences in the VQEG Phase I test dataset given by PSNR, M-SVD/Y Component, M-SVD/YCbCr and M-SVD/Edge Detection

every point represents one of 160 test video sequences. The vertical axis indicates the subject measurement denoted by DMOS while the horizontal axis is the objective measure output.

#### 4. CONCLUSION AND DISCUSSION

We presented a new objective video quality assessment system based on singular value decomposition. The graphical measure displays the amount of distortion as well as the distribution of error in all the frames of the tested video sequence. The numerical measure expressed as one single value is well correlated with subjective evaluation. Experiments on VQEG Phase I test imply the performance of M-SVD is considerably better than PSNR and other proponent models. In particular, the correlation of M-SVD model is improved by approximately 10% over PSNR in the first two metrics, and outperforms the SSIM by a clear margin as well.

It should be noted that we tested three models to obtain the correlation with subjective evaluation. In the first model, namely M-SVD/Y, only the luminance Y component was used in the computations. Experiments show that simply computing the difference of the original and the distorted frames on the Y component using equations 1, 5 and 7 provides reasonably good results comparable to other approaches. In the second model, denoted by M-SVD/YCbCr, the frame level quality error measure is a weighted summation of quality error indexes of Y, Cb and Cr components, each making a contribution with a weighting factor 0.8, 0.1 and 0.1, respectively. Equations 1, 5, 6 and 7 are employed in the second model whose performance is slightly better than the first model in the evaluation of metrics 2, 3 and 4. This implies the Human Visual System (HVS) is much more sensitive to the sharpness of the luminance component than those of the chrominance components. However, the chrominance components also contribute to the perceived quality error measurement. In the third model, namely M-SVD/Edge Detection, the difference of the edges of the luminance component is broadened with the respect to the characteristics of HVS. Model 3 has a better correlation with the perceived video quality in comparison with the first and the second models.

Our system is based on singular value decomposition (SVD). The computation of SVD is of order  $O(n^3)$ , which makes the computations slower for larger frame sizes. If the frame is segmented into smaller blocks as we use 8x8 for the block size, and the SVD is applied to each block, the total processing time is much lower. We also noticed randomly selected sample blocks in sample frames rather than using all blocks in frames of a video sequence would reduce the computation cost significantly, while still maintaining reasonably good results.

It should be mentioned that we found that M-SVD has relatively poor correlation with perceived video quality in such sequences of large global motions. This may be due to the designed system that does not incorporate motion information as one aspect of evaluation. In our experiments, we applied an approach to detect the occurrence of big motions and tried to improve the error measure in the sequences with motion. The issues relating to global motion are still under investigation.

#### REFERENCES

1. A. M. Eskicioglu and P. S. Fisher, "A survey of image quality measures for gray scale image compression," Proceedings of 1993 Space and Earth Science Data Compression Workshop, pp. 49-61, Snowbird, UT, April 2, 1993
2. J. O. Limb, "Distortion Criteria of the Human Viewer," IEEE Transactions on Systems, Man, and Cybernetics, Vol. 9, No. 12, pp. 778-793, December 1979.
3. J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," IEEE Transactions on Information Theory, Vol. 20, No. 4, pp. 525-536, July 1974.
4. D. J. Sakrison, "On the role of the observer and a distortion measure in image transmission," IEEE Transactions on Communications, Vol. 25, No. 11, November 1977.
5. C. F. Hall, "Subjective evaluation of a perceptual quality metric," Proceedings of SPIE, Vol. 310, pp. 200-204, 1981.
6. D. J. Granrath, "The role of human visual models in image processing," Proceedings of the IEEE, Vol. 69, No. 5, May 1981.
7. A. M. Eskicioglu and P. S. Fisher, "Image Quality Measures and Their Performance," IEEE Transactions on Communications, Vol. 43, pp. 2959-2965, December 1995.
8. D. Van der Weken, M. Nachtgael and E. E. Kerre, "A new similarity measure for image processing," Journal of Computational Methods in Sciences and Engineering, Vol. 3, No. 2, pp. 209-222, 2003.

9. A. Beghdadi and B. Pesquet-Popescu, "A new image distortion measure based on wavelet decomposition," 7th International Symposium on Signal Processing and Its Applications, Paris, France, July 1-4, 2003.
10. A. C. Bovik and S. Liu, "DCT-domain blind measurement of blocking artifacts in DCT-coded images," Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Salt Lake City, UT, May 7-11, 2001.
11. Z. Wang, A. C. Bovik and B. L. Evans, "Blind measurement of blocking artifacts in images," Proceedings of IEEE 2000 International Conferencing on Image Processing, Vancouver, BC, Canada, September 10-13, 2000.
12. Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," Proceedings of IEEE 2002 International Conferencing on Image Processing, Rochester, NY, September 22-25, 2002.
13. P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi, "A no-reference perceptual blur metric," IEEE 2002 International Conference on Image Processing, Rochester, NY, September 22-25, 2002.
14. E. -P. Ong, W. Lin, Lu, Z. Yang, S. Yao, F. Pan, L. Jiang and F. Moschetti, "A no-reference quality metric for measuring image blur," 7th International Symposium on Signal Processing and Its Applications, Paris, France, July 1-4, 2003
15. M. Carnec, P. Le Callet and D. Barba, "An image quality assessment method based on perception of structural information," 2003 International Conference on Image Processing, Barcelona, Spain, September 14-17, 2003.
16. A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "A multidimensional image quality measure using singular value decomposition," Proceedings of SPIE Image Quality and System Performance, Vol. 5294, pp. 82-92, San Jose, CA, January 19-20, 2004.
17. Z. Wang and A. Bovik, "A universal image quality index," IEEE Signal Processing Letters, Vol. 9, No. 3, pp. 81-84, March 2002.
18. Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error measurement to structural similarity," IEEE Transactions on Image Processing, Vol. 13, No. 4, April 2004.
19. A. M. Rohaly, J. Libert, P. Corriveau, and A. Webster, Editors, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," March 2000. Available at <http://www.vqeg.org/>.
20. P. Corriveau, et al., "Video Quality Experts Group: Current Results and Future Directions," Proc. SPIE Visual Communications and Image Processing, vol. 4067, June 2000.
21. A. Shnayderman, A. Gusev, and A. M. Eskicioglu, "An SVD-Based Gray-Scale Image Quality Measure for Local and Global Assessment," IEEE Transactions on Image Processing, February 2006.
22. A. Shnayderman, and A. M. Eskicioglu, "Assessment of Full-Color Image Quality With Singular Value Decomposition," IS&T/SPIE's 17th Symposium on Electronic Imaging, Image Quality and System Performance II Conference, San Jose, CA, January 16-20, 2005.
23. D. Kahaner, C. Moler and S. Nash, Numerical Methods and Software, Prentice-Hall, Inc, 1989.
24. S. O. Aase, J. H. Husoy and P. Waldemar, "A critique of SVD-based image coding systems," 1999 IEEE International Symposium on Circuits and Systems VLSI, Vol. 4, pp. 13-16, Orlando, FL, May 1999.
25. Z. Wang, L. Lu, and A. C. Bovik, "Video Quality Assessment Based on Structural Distortion Measure," Signal Processing: Image Communication, Vol. 19, No. 2, pp. 121-132, February 2004